The Journal of Abnormal
and
Social Psychology

Editor: M. BREWSTER SMITH, *New York University*
Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

DANIEL J. LEVINSON
Harvard Medical School

GARDNER LINDZEY
University of Minnesota

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN
Circulation Manager: HELEN ORR
Editorial Assistant: BARBARA CUMMINGS

VOLUME 56
1958

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1958 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 56

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

PAGE

Adelson, J. See Douvan, Elizabeth.
Ainsworth, L. H. **Rigidity, Insecurity, and Stress** 67
Alfert, Elizabeth. **Two Components of Assumed Similarity** 135
Anderson, J. K. See Miner, J. B.
Baer, D. M. See Gewirtz, J. L.
Baker, C. T. See Kagan, J.
Baker, R. W. See Vogel, W.
Bass, B. M., Pryer, Margaret W., Gaier, E. L., & Flint, A. W. **Interacting Effects of Control, Motivation, Group Practice, and Problem Difficulty on Attempted Leadership** 352
Beloff, Halla. **Two Forms of Social Conformity: Acquiescence and Conventionality** 99
Birney, R. C. **The Achievement Motive and Task Performance: A Replication** 133
Blackburn, H. L. **Effects of Motivating Instructions on Reaction Time in Cerebral Disease** 359
Blake, R. R. See Kimbrell, D. L.
Bobbitt, R. A. **The Repression Hypothesis Studied in a Situation of Hypnotically Induced Conflict** 204
Brehm, J. W. See Cohen, A. R.
Briskin, G. J. **Identification in Group Therapy** 195
Bugental, Daphne E., & Lehner, G. F. J. **Accuracy of Self-Perception and Group-Perception as as Related to Two Leadership Roles** 396
Burdock, E. I., Sutton, S., & Zubin, J. **Personality and Psychopathology** 18
Carey, G. L. **Sex Differences in Problem-Solving Performance as a Function of Attitude Differences** 256
Chapman, L. J. **Intrusion of Associative Responses Into Schizophrenic Conceptual Performance** 374
Christie, R., Havel, Joan, & Seidenberg, B. **Is the F Scale Irreversible?** 143
Cohen, A. R., Brehm, J. W., & Fleming, W. H. **Attitude Change and Justification for Compliance** 276
Coventry, J. See Taft, R.
Cowen, E. L., & Obrist, P. A. **Perceptual Reactivity to Threat and Neutral Words Under Varying Experimental Conditions** 305
Crosby, C. See Lambert, W. E.
Danielson, W. See Schramm, W.
Dibner, A. S. **Ambiguity and Anxiety** 165
Dickinson, J. See McArthur, C.
Dodge, Joan S. See Fiedler, F. E.
Doob, L. W. **Behavior and Grammatical Style** 398
Douvan, Elizabeth, & Adelson, J. **The Psychodynamics of Social Mobility in Adolescent Boys** 31
Errata 284
Faigin, Helen. **Social Behavior of Young Children in the *Kibbutz*** 117
Farber, M. L. **Anality, Political Aggression, Acquiescence, and Questionnaire Construction** 278
Ferguson, Eva D. **The Effect of Sibling Competition and Alliance on Level of Aspiration, Expectation, and Performance** 213
Fiedler, F. E., Dodge, Joan S., Jones, R.E., & Hutchins, E. B. **Interrelations Among Measures of Personality Adjustment in Nonclinical Populations** 345
Fisher, S., & Lubin, A. **Distance as a Determinant of Influence in a Two-Person Serial Interaction Situation** 230
Fitzgerald, B. J. **Some Relationships Among Projective Test, Interview, and Sociometric Measures of Dependent Behavior** 199
Fleming, W. H. See Cohen, A. R.
Flint, A. W. See Bass, B. M.
Foa, Uriel, G. **Empathy or Behavioral Transparency?** 62
French, Elizabeth G., & Thomas F. H. **The Relation of Achievement Motivation to Problem-Solving Effectiveness** 45
Gaier, E. L. See Bass, B. M.
Gewirtz, J. L., & Baer, D. M. **The Effect of Brief Social Deprivation on Behaviors for a Social Reinforcer** 49
Ginsberg, Rose. See Kenny, D. T.
Havel, Joan. See Christie, R.
Havelka, J. See Lambert, W. E.
Helper, M. M. **Parental Evaluations of Children and Children's Self-Evaluations** 190
Hoffman, M. L., Mitsos, S. B., & Protz, R. E. **Achievement Striving, Social Class, and Test Anxiety** 401
Holt, R. R. **Clinical *and* Statistical Prediction: A Reformulation and Some New Data** 1
Howe, E. S. **GSR Conditioning in Anxiety States, Normals, and Chronic Functional Schizophrenic Subjects** 183
Hutchins, E. B. See Fiedler, F. E.

PAGE
JACKSON, D. N. Independence and Resistance to Perceptual Field Forces 279
JESSOR, R. See ROBERTS, A. H.
JONES, R. E. See FIEDLER, F. E.
JOURARD, S. M., & LASAKOW, P. Some Factors in Self-Disclosure 91
KAGAN, J., SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Personality and IQ Change 261
KENNY, D. T., & GINSBERG, ROSE. The Specificity of Intolerance of Ambiguity Measures 300
KERLINGER, F. N. On Authoritarianism and Acquiescence: An Added Note to Bass and Messick and Jackson 141
KIDD, J. S. Social Influence Phenomena in a Task-Oriented Group Situation 13
KIMBRELL, D. L., & BLAKE, R. R. Motivational Factors in the Violation of a Prohibition 132
KING, G. F. Differential Autonomic Responsiveness in the Process-Reactive Classification of Schizophrenia 160
KNAPP, R. H. See MCCLELLAND, D. C.
KOGAN, N., & TAGIURI, R. Interpersonal Preference and Cognitive Organization 113
KREBS, A. M. Two Determinants of Conformity: Age of Independence Training and Achievement 130
LAMBERT, W. E., HAVELKA, J., & CROSBY, C. The Influence of Language-Acquisition Contexts on Bilingualism 239
LAMBERT, W. W. See TRIANDIS, H. C.
LASAKOW, P. See JOURARD, S. M.
LAZARUS, R. S. See VOGEL, W.
LEHNER, G. F. J. See BUGENTAL, DAPHNE E.
LESSER. G. S. Extrapunitiveness and Ethnic Attitude 281
LUBIN, A. See FISHER, S.
LUBORSKY, L. See SHEVRIN, H.
MCARTHUR, C., WALDRON, ELLEN, & DICKINSON, J. The Psychology of Smoking 267
MCCLELLAND, D. C., STURR, J. F., KNAPP, R. H., & WENDT, H. W. Obligations to Self and Society in the United States and Germany 245
MCGINNIES, E. See MITNICK, L. L.
MANDLER, G., MANDLER, JEAN M., & UVILLER, ELLEN T. Autonomic Feedback: The Perception of Autonomic Activity 367
MANDLER, JEAN M. See MANDLER, G.
MATARAZZO, J. D. See MATARAZZO, RUTH G.
MATARAZZO, RUTH G., MATARAZZO, J. D., SASLOW, G., & PHILLIPS, JEANNE S. Psychological Test and Organizmic Correlates of Interview Interaction Patterns 329
MEISSNER, W. W. Affective Response to Psychoanalytic Death Symbols 295
MILHOLLAND, J. E. The Scale Values of Statements of Attitude Toward the Law After a 25-Year Interval 131
MINER, J. B., & ANDERSON, J. K. Intelligence and Emotional Disturbance: Evidence From Army and Veterans Administration Records 75
MISCHEL, W. Preference for Delayed Reinforcement: An Experimental Study of a Cultural Observation 57
MITNICK, L. L., & MCGINNIES, E. Influencing Ethnocentrism in Small Discussion Groups Through a Film Communication 82
MITSOS, S. B See HOFFMAN, M. L.
NAKAMURA, C. Y. Conformity and Problem Solving 315
NELSON, VIRGINIA L. See KAGAN, J.
OBRIST, P. A. See COWEN, E. L.
PHILLIPS, JEANNE S. See MATARAZZO, RUTH G.
PROTZ, R. E. See HOFFMAN, M. L.
PRYER, MARGARET W. See BASS, B. M.
ROBERTS, A. H., & JESSOR, R. Authoritarianism, Punitiveness, and Perceived Social Status 311
SARASON, I. G. Interrelationships Among Individual Difference Variables, Behavior in Psychotherapy, and Verbal Conditioning 339
SASLOW, G. See MATARAZZO, RUTH G.
SCHRAMM, W., & DANIELSON, W. Anticipated Audiences as Determinants of Recall 282
SEIDENBERG, B. See CHRISTIE, R.
SHEVRIN, H., & LUBORSKY, L. The Measurement of Preconscious Perception in Dreams and Images: An Investigation of the Poetzl Phenomenon 285
SONTAG, L. W. See KAGAN, J.
STOTLAND, E., & ZANDER, A. Effects of Public and Private Failure on Self-Evaluation 223
STURR, J. F. See MCCLELLAND, D. C.
SUTTON, S. See BURDOCK, E. I.
TAFT, R., & COVENTRY, J. Neuroticism, Extraversion, and the Perception of the Vertical 139
TAGIURI, R. See KOGAN, N.
THOMAS, F. H. See FRENCH, ELIZABETH G.

PAGE
TRANKELL, A. Was Lars Sexually Assaulted? A Study in the Reliability of Witnesses and of Experts. 385
TRIANDIS, H. C., & LAMBERT, W. W. A Restatement and Test of Schlosberg's Theory of Emotion With Two Kinds of Subjects From Greece. 321
UVILLER, ELLEN T. See MANDLER, G.
VOGEL, W., BAKER, R. W., & LAZARUS, R. S. The Role of Motivation in Psychological Stress. 105
WALDRON, ELLEN. See MCARTHUR, C.
WEISS, W. The Relationship Between Judgments of a Communicator's Position and Extent of Opinion Change. 380
WENDT, H. W. See MCCLELLAND, D. C.
YATES, A. J. The Application of Learning Theory to the Treatment of Tics. 175
ZANDER, A. See STOTLAND, E.
ZUBIN, J. See BURDOCK, E. I.

CLINICAL *AND* STATISTICAL PREDICTION: A REFORMULATION AND SOME NEW DATA[1]

ROBERT R. HOLT

Research Center for Mental Health, New York University

THE controversial discussions started a few years ago by Meehl's tightly packed little book, *Clinical vs. Statistical Prediction* (10), still continue—especially among graduate students in psychology, most of whom have to read it. Clinical students in particular complain of a vague feeling that a fast one has been put over on them, that under a great show of objectivity, or at least bipartisanship, Professor Meehl has actually sold the clinical approach up the river. The specific complaints they lodge against the book are, in my opinion, mostly based on misinterpretations, wishful thinking, or other errors, yet I have felt for some time that there was something valid in the irrational reaction without knowing why.

What I propose to show here is that clinicians do have a kind of justified grievance against Meehl, growing out of his formulation of the issues rather than his arguments, which are sound. Finally, I want to offer a slightly different approach to the underlying problems, illustrated by some data. It may not quite make the lion lie down with the lamb, but I hope that it will help us all get on with our business, which is the making of a good science and profession.

THE ISSUES RESTATED

Meehl's book contains a review of the controversy, a logical analysis of the nature of clinical judgment, a survey of empirical studies, and some conclusions. I am not going to go into his treatment of the logical issues and his psychological reconstruction of clinical thinking; for the most part, I agree with this part of the book and consider it a useful contribution to methodology. I want to focus rather on his conception of what the issues are in the controversy, on his treatment of the evidence, and on some of his conclusions.

Many issues make better reading when formulated as battles, and the field of the assessment and prediction of human behavior has not lacked for controversy-loving gauntlet-flingers. The sane and thoughtful voices of Horst and his collaborators, urging compromise and collaboration (5), have been shouted down by the warcries of such partisans as Sarbin (14) on the actuarial side and Murray (12) on the clinical or (as he put it) organismic. Meehl approached the problem with a full awareness of the feelings on both sides, and apparently with the hope that the therapeutic ploy of bringing them all out into the open at the beginning would enable him to discuss the issues objectively.

In a recent discussion of the stir his book has raised, Meehl has expressed surprise and dismay (11) that his effort to take a balanced and qualified position has led so many people to misunderstand him as claiming that clinical prediction has been proved worthless. Yet he is not blameless; by posing the question of clinical *vs.* statistical prediction, he has encouraged two warring camps to form. This in turn makes it appear all the more compellingly that there *are* two clear-cut types of prediction to be compared.

The root difficulty, I believe, lies in Meehl's acceptance of *clinical* and *actuarial* as concepts that can without further analysis be meaningfully applied to a variety of predictive endeavors of an experimental or practical sort. Accepting them as valid types, he can hardly do anything other than pit one against the other and try to decide what is the proper sphere of exercise for each. But the terms in this antithesis mean many things; they are constellations of parts that are not perfectly correlated and can be separated.

The issue cannot therefore be sharply drawn so long as we speak about anything as complex as "clinical prediction" or "the clinical method." Rather, I think the central issue is the *role of clinical judgment in predicting human behavior*. By clinical judgment here, I mean nothing more complicated than the problem-solving or decision-reaching behavior of a person who tries to reach conclusions on the basis of facts and theories already available to him by thinking them over.

---

[1] The research reported in this paper was supported by the Veterans' Administration, the New York Foundation, and the Menninger Foundation, and was carried out jointly by Dr. Lester Luborsky and the author, with the collaboration of Drs. Wm. R. Morrow, David Rapaport, and S. K. Escalona.

Let us make a fresh start, therefore, by examining the logical structure of the predictive process with an eye to locating the points where clinical judgment may enter. The following five-step process is idealized, and in practice some of the steps are more or less elided, but that does not hurt this analysis.

First, if we are to predict some kind of behavior, it is presupposed that we acquaint ourselves with what we are trying to predict. This may be called job analysis or the study of the criterion. Perhaps those terms sound a little fancy when their referent is something that seems so obvious to common sense. Nevertheless, it is surprising how often people expend a great deal of time and effort trying to predict a kind of behavior about which they know very little and apparently without even thinking that it might help if they could find out more. Consider the job of predicting outcome of flight training, for example. Many attempts to predict passing or washing out have been made by clinicians without any direct experience in learning to fly, without any study of flight trainees to see what they have to *do* in order to learn how to fly a plane, or of the ways they can fail to make the grade (cf. 4).

There is a hidden trick in predicting something like success in flight training, because that is not itself a form of behavior. It is an outcome, a judgment passed by someone on a great deal of concrete behavior. The same is true for grades in college, success in any type of treatment, and a host of other criteria that are the targets in most predictive studies. Because it is hidden by the label, there is a temptation to forget that the behavior you should be trying to predict exists and must be studied if it is to be rationally forecast. In the highly effective pilot selection work carried out by psychologists during the war, careful job analyses were an important step in the total predictive process and undoubtedly contributed a good deal to the over-all success.

This first stage is hardly a good point at which to try to rely on clinical judgment. The result is most likely to be that guesses, easy and arbitrary assumptions, and speculative extrapolations will attempt to substitute for real information. And no matter how remarkable clinical judgment may sometimes be, it can never create information where there is none.

The *second* logical step is to decide what intervening variables need to be considered if the behavior is to be predicted. As soon as we get away from the simplest kind of prediction—that someone will continue to act the way he has been acting, or that test behavior A will continue to correlate (for an unknown reason) with criterion behavior B—we have to deal with the inner constructs that mediate behavior and the determining situational variables as well. You cannot make a rational choice of the kind of information you will need to have about a person to make predictions without some intervening variables, though they may remain entirely implicit. At this point, judgment enters—always, I think, though it may be assisted by empirical study. The best practice seems to be to give explicit consideration to this step, and to supply judgment with as many relevant facts as possible. This means studying known instances, comparing people who showed the behavior in question with others who in the same situation failed to.

All too often, when the problem of intervening variables is considered at all, it is handled by unaided clinical judgment. For example, in the Michigan project on the selection of clinical psychologists (7), a good many personality variables were rated, but there was no previous work highlighting the ones that might be related to success as a clinical psychologist. It was left up to each judge to form his own conception (from experience, theory, and guess) about what qualities mattered most. Again, this puts a greater burden on clinical judgment than it should reasonably be asked to bear. Yet some clinicians seem to have the mistaken notion that they are being false to their professional ideals if they stir from their armchairs at this point; nothing could be further from the best in clinical tradition, which is unashamedly empirical.

Third, it is necessary to find out what types of data afford measures or indications of the intervening variables, and can thus be used to predict the criterion behavior. If a good job has been done of the preceding step, it may be possible to rely entirely on judgment to make the preliminary selection of appropriate means of gathering predictive data. For example, if a job analysis and study of persons who have done well at the performance in question both suggest that verbal intelligence and information of a certain type are the main requisites, it would be easy to make good guesses about appropriate instruments to provide the predictive data. I use the word "guesses" deliberately, however, to emphasize the fact that judgment can do no more than supply hypotheses; it cannot substitute for an empirical trial to see whether in fact and under the conditions of this particular study the likely-looking instruments do yield data that predict the criterion.

Notice that almost any actuarial predictive system presupposes carrying through this step. If there is to be an actuarial table, one has to collect great numbers of cases to determine the success frequencies for each cell in the table; if a regression equation is to be used, there must be a preliminary study to fix the beta weights. Unfortunately, it is possible to work clinically *without* first getting an empirical check on one's hypotheses about likely-seeming instruments. At the risk of boring you, I repeat: there simply is no substitute for empirical study of the actual association between a type of predictive data and the criterion. Just as judgment is indispensable in forming hypotheses, it cannot be used to test them.

Perhaps this caution seems misplaced. Do I seem to be urging that you should first *do* a predictive study before embarking on one? I am. That is exactly what happens in actuarial prediction: the formula or table being pitted against judgmental prediction is typically being *cross*-validated, while in none of the studies Meehl cites were the clinical predictions under test being cross-validated. This alone is a major reason to expect superior performance from the actuarial predictions, and again it is a disadvantage under which the clinician by no means has to labor.

The next step, the *fourth* one, is to gather and process the data to give measures of the intervening variables. Meehl clearly recognizes that at this point clinical judgment either may play a large role or may be minimized. At one extreme, the data-yielding instrument may be a machine, a gadget like a complex coordination tester, which automatically counts errors and successes and makes a cumulative record of them. The resulting numbers may be directly usable in a regression equation without the intervention of anyone more skilled than a clerk. At an intermediate level, scoring most psychological tests requires a modicum of clinical judgment, though a high degree of reliability may be attained. At the other extreme is the interview; a great deal of clinical judgment is needed to convert the raw data into indices of the constructs the interviewer wants to assess.

It is easily overlooked that judgment needs the help of empirical study in this phase of the work too. The clinician's training supplies this empirical base in large part, but when he is using a familiar instrument to measure unusual intervening variables, or when he is working with an unfamiliar instrument, judgment grows fallible, and it is no more than prudent to piece it out by careful study of the same kind of predictive data on known subjects on whom the intervening variables have been well assessed independently.

The *fifth* and final step is the crucial one: at last the processed predictive data are *combined* so as to yield definite predictions in each case. The job can be done by clinical judgment, or it can be done by following a fixed rule (an actuarial table or regression equation) in a mechanical way. That much is clear; indeed, this is the locus of Meehl's main interest. I am taking it as granted that a clinician often integrates data in a different way than a statistician—as Meehl says, by performing a creative act, constructing a model of the person from the given facts put together with his theoretical understanding and thus generating perhaps a new type of prediction from a pattern he has never encountered before. We are all curious to know how well good clinicians can do it, and wonder if actuarial combination of data can do as well or better.

But it now seems plain that Meehl has been *too much* interested in this last stage, and as a result has neglected to pay enough attention to the way the earlier aspects of the predictive process were handled in the studies he has reviewed. Here I want to state my main critical point: *If two attempts to predict the same behavior differ significantly in the role played by clinical judgment as against actual study of the facts in one or more of the four earlier parts of the predictive process, a comparison of the suc-*

cesses of the two attempts can tell us nothing definite about the effectiveness of clinical judgment at the final, crucial stage. For this reason, in none of the 20 studies Meehl cites were the comparisons pertinent to the point. Particularly at the vital third step, the predicting statisticians have had the advantage of having previously studied the way their predictive data are related to the criterion; the clinicians have not.

If your reaction is, "So much the worse for the clinicians; nobody stopped them," I am afraid you are thinking about a different question from the one Meehl has raised. If the issue were whether some clinicians have made themselves look foolish by claiming too much, then I should agree: these studies show that they have, and unhappily, they have brought discredit on clinical methods generally. But the studies cited by Meehl and more recently by Cronbach in the *Annual Review of Psychology* (2) unfortunately have too many flaws at other points to tell us what clinical judgment can or cannot do as a way of combining data to make predictions. It is as if two riflemen were having a target match, but one took a wrong turn on the way to the shoot, never showed up, and lost by default. He demonstrated himself to be a poor driver, perhaps, but we never found out how well he could shoot, which is what we really wanted to know.

The other point I want to make in connection with the five-step analysis of the predictive process is this: Since there are so many ways in which clinical judgment can enter, for better or for worse, it makes little sense to classify every attempt to predict behavior on one side or the other of a simple dichotomy, clinical vs. statistical. There can be many types of clinical and actuarial combinations, and many are in fact found in Meehl's mixed bag.

For purposes of exposition, I should like to suggest an only slightly extended typology. Extracting from the best actuarial studies those parts of their procedure during the first four steps that are simply the application of common sense and the scientific method, I propose that we make it quite plain that these can be separated from actuarial prediction at the final step by creating a third type. Thus we should have:

Type I. *Pure actuarial:* Only objective data are used to predict a clear-cut criterion by means of statistical processes. The role of judgment is held to a minimum, and maximal use is made of a sequence of steps exemplified in the most successful Air Force studies in selecting air crew personnel (job analysis, item analysis, cross-validation, etc.).

Type II. *Naive clinical:* The data used are primarily qualitative with no attempt at objectification; their processing is entirely a clinical and intuitive matter, and there is no prior study of the criterion or of the possible relation of the predictive data to it. Clinical judgment is at every step relied on not only as a way of integrating data to produce predictions, but also as an alternative to acquaintance with the facts.

Type III. *Sophisticated clinical:* Qualitative data from such sources as interviews, life histories, and projective techniques are used as well as objective facts and scores, but as much as possible of objectivity, organization, and scientific method are introduced into the planning, the gathering of the data, and their analysis. All the refinements of design that the actuarial tradition has furnished are employed, including job analysis, pilot studies, item analysis, and successive cross-validations. Quantification and statistics are used wherever helpful, but the clinician himself is retained as one of the prime instruments, with an effort to make him as reliable and valid a data-processor as possible; and he makes the final organization of the data to yield a set of predictions tailored to each individual case.

If we now re-examined the studies cited by Meehl and Cronbach, we see that most of them have pitted approximations to Type I actuarial predictive designs against essentially Type II naive clinical approaches. It seems hardly remarkable that Type I has generally given better results than Type II; indeed, the wonder should be that the naive clinical predictions have done as well as they have, in a number of instances approaching the predictive efficiency of actuarial systems.

Other studies cited have come closer to comparing Type II with Type III—naive vs. sophisticated clinical prediction instead of clinical vs. statistical. For example, the prognostic studies by Wittman (16) compared predictions of reaction to shock treatment made in a global way at staff conference with a system she devised. But her system used highly judgmental predictive variables, as Meehl himself points out (ranging from *duration of psychosis* to *anal erotic vs. oral erotic*), and they were combined using a set of weights assigned on judgmental, not statistical grounds.[2] What she showed was that a systematic and comprehensive evaluation of the thirty items in her scale (all based on previous empirical work) made better predictions of the outcome of shock treatment than global clinical judgments not so organized and guided. A study of movement in family case work by Blenkner came to a very similar conclusion with somewhat different subject matter (1). When social workers rated an initial interview according to their general impressions, they were unable to predict the outcome of the case, whereas when their judgments were organized and guided by means of an outline calling for appraisals of five factors which had been shown in previous studies to be *meaningfully*, not statistically related to the criterion, then these judgmentally derived predictive variables, combined (like Wittman's) in an a priori formula, predicted the criterion quite well. Yet both studies are tallied as proving actuarial predictions superior to clinical.

---

[2] It is true that the weights were applied in the same way for all cases; in this respect, the system deviates from the ideal Type III.

Meehl's conclusion from his review of this "evidence" is that clinical prediction is an expensive and inefficient substitute for the actuarial method, and one that keeps clinicians from using their talents more constructively in psychotherapy or exploratory research.

The evidence available tells us hardly anything about the relative efficacy of clinical judgment in making predictions. The weight of numbers should not impress us; as long as the studies don't really bear on the issue, no matter how many are marshalled, they still have no weight. Remember the *Literary Digest* poll: many times more straw votes than Gallup used, but a faulty sampling principle, so that piling up numbers made the conclusion less valid as it got more reliable. Moreover, the studies tallied are so different in method, involving varying amounts of clinical judgment at different points (in the "actuarial" instances as well as the "clinical" ones), that they cannot sensibly be added together.

What is fair to conclude, I think, is that many clinicians are wasting their time when they try to fall back on their clinical judgment in place of knowing what they are talking about. They have been guilty of over-extending themselves, trying to predict things they know nothing about, and learning nothing in the process of taking part in what Cronbach calls "horserace experimental designs," in which clinicians and statisticians merely try to outsmart each other. A multiplication of such studies will not advance clinical psychology.

One kind of comparative study might teach us something even though it would be hard to do properly: simultaneous attempts to predict the same criterion from the same data by clinicians and statisticians *who have gone through the same preliminary steps*. As the statistician studies the original group to determine the critical scores for his multiple cutting point formula (or whatever), the clinician will study the configurations of these scores in individuals of known performance. Then we will see how their respective predictive techniques work.

Does it really make sense, however, for both to use the same data and predict the same criterion? A second possibility would be for two otherwise equally sophisticated methods to predict the same criterion, each using the kind of data most appropriate to it. Or, third, the more clinical and the more statistical methods would not predict the same criterion, but each would undertake to predict the kind of behavior it is best suited to, using the most appropriate kind of data.

Doesn't this third proposal abandon experimental controls necessary for intelligible results? To some extent, yes; but one may have to give up some control to avoid absurdity. As long as clinician and statistician are trying to predict the same criterion, the clinician is likely to be working under a severe, though concealed, handicap. The study will usually have been designed by the statistician; that is his business. He will naturally choose the kind of criterion for which *his* methods seem best adapted; indeed, the nature of his method makes it impossible for him to choose the kind of predictive task that would be most congenial to the clinician, such as drawing a multidimensional diagnostic picture of a total personality, or predicting what a patient will do next in psychotherapy. Thus, the statistician takes advantage of the foolish boast of the clinician, "Anything you can do, I can do better," and plans the contest on his own grounds. The clinician ends up trying to predict grade-point average in the freshman year by a "clinical synthesis" of high school grades and an intelligence test. This is a manifest absurdity; under the circumstances, how could the clinician do other than operate

Research
DAVID HA COLLEGE

like a second-rate Hollerith machine? If clinical judgment is really to be tested, it must operate on data that are capable of yielding insights. Moreover, it makes hardly any more sense to expect it to grind out numerical averages of course grades than to expect an actuarial table to interpret dreams.

For reasons of this kind, McArthur (9) recently called for studies of the third type just listed, and maintained that there have as yet been no studies in which the clinician has been given a chance to show what he can do on his own terms. I want therefore to present briefly some results from one such attempt: a study in which clinicians tried to predict criteria of their own choosing, using clinical types of data—interviews and psychological tests (mainly projective techniques).[3] Since some preliminary reports of this work have already been quoted as showing the ineffectiveness of clinical predictions (cf. 2), I have a special desire to set the record straight.

VALIDATING NAIVE AND SOPHISTICATED CLINICAL PREDICTIONS: SOME NEW DATA

The project was an effort to improve the methods by which medical men were selected for specialty training in the Menninger School of Psychiatry. It was begun by Dr. David Rapaport, together with Drs. Karl A. Menninger, Robert P. Knight, and other psychiatrists at the Menninger Foundation at the time the Menninger School of Psychiatry was founded 11 years ago. In the late summer of 1947, Dr. Lester Luborsky and I began work on it, and since then we have jointly carried major responsibility for the project although quite a number of other people have made important contributions.

Our work consisted of two predictive studies. Following the terminology suggested above, one used a naive clinical method, while the other was an attempt at a more sophisticated clinical method. The naive clinical design was simple: Psychiatrists and psychologists used their favorite means of assessing personality to forecast the level of future performance of psychiatric residents at the time when they applied to the School. The applicant came to Topeka after some preliminary correspondence, having survived a rough screening of credentials and application forms. He was seen by three psychiatrists, each of whom interviewed him for about an hour, rated his probable success as a resident on a 10-point scale, and made a recommendation: Take, Reject, or Doubtful. The psychologist made similar ratings and recommendations after analyzing results of a Wechsler-Bellevue, Rorschach, and Word Association test. In addition, both psychologists and psychiatrists submitted brief qualitative reports of their appraisals of the man's positive and negative potentialities. All of the data and predictions were turned over to the Admissions Committee, which made the final decision to accept or reject each man.

During the years of the project, from 1946 through 1952, six successive classes of residents were chosen. The first 456 applicants who went through this procedure formed our experimental population (excluding small numbers of Negroes, women, and persons from Latin-American and non-European cultures, since these minorities offered special problems of assessment). A little over 62 per cent of these applicants were accepted by the Committee, but only 238 actually entered the School; 46 changed their minds and went elsewhere. Nevertheless, we kept in touch with them and the 172 rejectees by a mail follow-up questionnaire for several years, so that we have data on certain aspects of their subsequent careers.

The clinicians making the predictions had in some cases had considerable experience in training psychiatric residents, but there was no explicit job analysis or preliminary study of criterion groups. They simply fell to and made their predictions and their decisions.

To test the validity of these clinical decisions, let us use as a criterion, first, whether or not a man passed the certifying examination of the American Board of Psychiatry and Neurology—the criterion set up by the specialty itself. We have this information on all subjects from the lists published by the Board. The Admissions Committee's decisions had a good deal of validity as predictors of this criterion: 71 per cent of the men they voted to accept had passed the Board examination in psychiatry by the end of 1956, while only 36 per cent of the rejected candidates had done so. This difference is significant at better than the .001 point. The recommendations made by interviewers (taking them all as a group) and the recommendations of the psychological testers to accept or reject likewise were highly valid predictors of this criterion, significances in both cases also being beyond .001.

---

[3] This study is presented at length in a forthcoming book (3).

It is interesting that the Committee decisions were slightly *better* at predicting both staying in psychiatry and passing the certification examination of the American Board of Psychiatry and Neurology—better than either the psychiatric interviewers or the psychological testers. It is possible, however, that much or all of this apparent superiority is due to the fact that rejection by the committee did discourage a few applicants from seeking training elsewhere.

Data of this kind are encouraging to the people trying to run a school of psychiatry but hard to interpret in a larger context. Who knows but that an actuarial table based on objectively ascertainable facts like grades in medical school, marital status, age, etc. might not have done just as well? We never took such a possibility very seriously, but we did try out a few such objective predictors in our spare time, just out of curiosity. None of them showed any particular promise as a predictor of any criterion taken alone, though it is possible that patterns of them such as an actuarial table uses might have operated a little better than chance.

The criteria on which we spent most time and labor were measures of competence in psychiatric work during the last two years of the three-year residency. Whenever a resident completed a period of time on a particular service, we would interview the staff men, consultants, and others who had directly supervised his clinical work, and get them to rate it quantitatively. The resulting criterion measure had a coefficient of internal consistency above **.9**[4] (for the last few classes), and we have every reason to think it has a great deal of intrinsic validity. We also got the residents to rate each other's work. The reliability of their pooled ratings of over-all competence is also **.9**, and this criterion (which we call "Peers' Evaluations") correlates from **.66** to **.78** with Supervisors' Evaluations.

These criterion judgments enable us to test the validity of the predictive *ratings*. The validity correlations are not exciting, though for the entire group of residents in the Menninger School of Psychiatry they are all significantly better than zero at the one per cent point. Taking the *mean* of the ratings given by the psychiatrists who interviewed an applicant, this predictor correlates **.24** with Supervisors' Evaluations. The predictions of the psychological testers were not significantly better: the validity coefficient is **.27.** There was some fluctuation from class to class, the interviewers' validities varying from exactly .00 to *.52*, the tester's validities from .12 to **.47.** Likewise, the validities of ratings made by individual clinicians vary over the same range: psychiatric interviewers from .01 to **.27** ($N = 93$) or *.47* (significant at only the 5 per cent point because the N was only 13), and psychologists from .20 to **.41** ($N = 40$). At the same time, those individual clinicians all did much better in making the basic discrimination: recommending acceptance of men who actually became psychiatrists.

These correlations are nothing to get excited about, and nothing to be ashamed of either, particularly in view of the restriction of range in this selected, accepted sample. They show that the naive clinical method depends a good deal on the ability of the particular clinician doing the predicting, and that—at least in this study—a pooling of judgment helped make up for the deficiencies of individuals.

Let us turn now to the second experimental design, which I have called a *sophisticated clinical* type of prediction. I shall have to skip lightly over many complicated details, and make things look a little more orderly than they actually were. The design included a job analysis of the work done by psychiatric residents, which was broken down into a few major functions (such as diagnosis, psychotherapy, administration of wards) and 14 more specific aspects of work. Then we attempted to specify attributes of personality that would facilitate or hinder a man in carrying out such work, first by collecting the opinions of persons who had had long experience in training psychiatrists, psychotherapists, or psychoanalysts. The second way we went about it was to make an intensive study of a dozen excellent residents and a dozen who were rated at the bottom by their supervisors. We went over all the original assessment data on them, interviewed them and tested them extensively, trying out many novel approaches and then seeing what discriminated these known extreme groups. Thus, we learned what personological constructs differentiated good from poor residents, and what tests and test indicators gave evidence of these constructs. Hoping to guide clinical judgment in the use of interviews and projective tests, we used the data from these small samples of extremes to help us write detailed manuals on the use of the interview, TAT, Rorschach, and other techniques in the selection of psychiatric residents. The manuals listed discriminating cues, both positive and negative, which were to be summed algebraically. We then made preliminary cross-validations of these manuals (as many as we could) with encouraging results (cf. 8) and revised them after studying predictive successes and failures.

---

[4] Correlations in bold type are significant at the one per cent point, those in italics, at the five per cent point. One-tailed tests are used throughout to test the null hypothesis that the predictor does not correlate *positively* with the criterion.

As a last step, we set up another predictive study to submit our manuals to a final cross-validation on a group of 64 subjects and to accomplish several other purposes at the same time. Four psychologists served as judges; each of them scored tests or interviews according to our manuals and also made free clinical judgments based on increasing amounts of data. Two of the judges made such predictive ratings after going through an entire file of assessment data: credentials, intellectual and projective tests, and a recorded and transcribed interview.

How did we make out? Considering first the manuals, only indifferently well. Of the six, two proved worthless (TAT Content and a special projective test); the other four all showed more or less promise, but there was none that yielded consistently significant validities regardless of who used it. Reliability, in terms of scorer agreement, was on the whole not very good, for a good deal of clinical judgment was still demanded. Consider, for example, one TAT cue that worked well for one judge (validity of *.26* against Supervisor's Evaluations of Overall Competence). This cue called for judgments of the *originality* of each TAT story, obviously a matter on which psychologists might fairly easily disagree: scores by Judges I and II correlated −.04. The validities attained by the manuals for the Interview, Rorschach, Formal Aspects of the TAT, and Self-Interpretation of the TAT were on about the same general level as those from our first, naive clinical design—mostly in the .20's.

Now for the free clinical predictive ratings. When the judge had only one test or an interview to go on he usually added little by going beyond the manual and drawing on his general experience and intuition. Some judges did slightly better with free ratings than with manual scores, some a little worse. At this point, you may wonder at all this exposition for so small a result: barely significant validity coefficients, about the same size as those from a naive clinical approach, despite the attempt to create a sophisticated clinical predictive system that involved many actuarial elements. I believe that the lesson of our findings up to this point is simple: *With an inadequate sample of information about a person, no matter how sophisticated the technique of prediction, there is a low ceiling on the predictive validity that can be attained.* In our experience, even a battery of two or three tests (if exclusively projective), or an interview and a couple of projective techniques, does not give an adequate informational sample to enable clinicians to make very accurate predictions of complex behavioral outcomes over a period of three years.

Look next at the results when experimental judges made their predictions from as complete a body of information about a man as could be assembled at the time he applied. The hard-headed statistical expectation *should* be that validities would at best remain at the same level, and more likely would decline. The widely read preliminary report of the Michigan project on the selection of clinical psychologists (6) reported declining validities as increasing amounts of information were made available to judges (in a design which, for all its complexity, was essentially a naive clinical one). Not so many people have read the full final report (7), where this issue is not discussed; it is necessary to pore over many tables and tally numbers of significant correlations at various stages to find out for oneself that with the final criteria there was a slight *rising* trend in validities of clinical predictions as the

TABLE 1

SOME VALIDITIES FROM SYSTEMATIC CLINICAL ASSESSMENT OF APPLICANTS FOR RESIDENCIES IN THE MENNINGER SCHOOL OF PSYCHIATRY

Predictors:	Validities of Predictors Against Criterion Evaluations of:									
	Overall Competence		Competence in Psychotherapy		Competence in Diagnosis		Competence in Management		Competence in Administration	
	Sup.[a]	Peer[a]	Sup.	Peer	Sup.	Peer	Sup.	Peer	Sup.	Peer
Predictive Ratings										
Judge I: PRT[b]	*.26*	.23	.12	*.26*	.13	.21	*.31*	.00	.20	.10
Judge II: TAT	−.10	−.02	−.16	−.01	−.05	.20	−.08	.01	.04	.11
Judge I: All data	**.57**	**.52**	**.48**	**.55**	**.58**	**.42**	**.52**	*.36*	**.55**	**.42**
Judge II: All data	*.22*	**.48**	.15	**.36**	*.24*	**.42**	.13	.24	*.24*	*.27*
Liking Ratings										
Judge I: PRT	*.29*	*.34*	.16	**.35**	*.24*	**.36**	.15	.17	.19	.16
Judge II: TAT	−.02	.13	−.08	.15	.00	*.30*	−.17	−.02	−.14	.10
Judge I: All data	**.58**	**.64**	**.45**	**.58**	**.51**	**.52**	**.52**	**.52**	**.50**	**.46**
Judge II: All data	*.25*	**.49**	.20	**.47**	*.21*	**.56**	.10	*.30*	.18	*.30*

[a] Sup. = Supervisors' Evaluations; Peer = Peers' (Sociometric) Evaluations.

[b] PRT = Picture Reaction Test, a specially devised projective test similar to TAT.

Numbers of cases: For Judge I—Supervisors' Evaluations, PRT: 63, all data: 37
Peers' Evaluations, PRT: 45, all data: 30
For Judge II—Supervisors' Evaluations, TAT: 63, all data: 64
Peers' Evaluations, TAT: 45, all data: 46

amount of information available to the predicting judge was increased.

The same thing is true of our results, but in a dramatic and unmistakable way (see Table 1). Considering only the two judges who went through the entire mass of material, their final free clinical ratings of Over-all Competence correlated **.57** and *.22* with Supervisors' Evaluations and **.52** and **.48** with the sociometric Peers' Evaluations. In considering these correlations, remember that they are attentuated by a significant restriction of range, since all subjects had successfully passed through an Admissions Committee screening which had considerable validity. They have not been corrected for less than perfect reliabilities, either.

An incidental finding is even more remarkable. The predictive analysis was made approximately a year after the assessment data had been gathered. The judges went right through the entire series of cases, making their ratings in a *blind* analysis; names and other identifying data were concealed, and there was no direct contact with the subjects. Nevertheless, judges formed rather vivid impressions from the material, including a feeling of how well they would like each candidate personally. For control purposes, they were required to rate this feeling of *liking*. When we undertook to correlate the liking rating with predictors and criteria so as to partial out this possible source of error, we found that it was the best predictor we had! These ratings of liking by Judges I and II correlated highly with their predictive ratings, but even more highly with Supervisors' Evaluations (**.58** and *.25*) and especially with Peers' Evaluations (**.64** and **.49**).[5] A study of these liking ratings suggests to us as the most plausible explanation that they differed from our intentional clinical predictions in being somewhat more irrational, affective—perhaps intuitive—reactions to the same data.

---

[5] One consequence of the delay between the gathering of the data and their analysis was that some of the *S*s who entered the Menninger School of Psychiatry became known to the predictive judges, raising the possibility of contamination of their predictions by criterion-relevant knowledge. Despite the fact that the analysis of the assessment material was done "blind," identifying data having been removed or concealed, Judge I fairly often recognized the identity of *S*s at the final stage of analysis. He therefore did not make predictive ratings, which is why his *N*s are so low for this stage. There were a few borderline instances of partial or questionable recognition, however, in which Judge I (four cases) or II (two cases) had some information or misinformation about the subject. If these cases are eliminated, Judge II's validities go up more often than they go down, the range being from a decrease of .16 to an increase of .15. Judge I's validities against Supervisors' Evaluations are negligibly affected, the range of effects extending from a loss of .08 to a gain of .20. Most of his validities against Peers' Evaluations were more seriously affected, however, especially at the final (all-data) stage of analysis, where losses of up to 28 correlation points occurred. On the whole, however, even a very conservative handling of the problem of possibly contaminated cases does not change the essential import of the results: It was still possible for Judge I to obtain four validities of **.50** or higher, and for Judge II to obtain two validities

In all of the correlations I have been citing, you will perhaps have noticed that one judge consistently did slightly better than the other. This is certainly to be expected. When clinical judgment is the main technique of processing data, there are bound to be differences due to the skill of clinicians in doing this particular job.

Are we justified in citing these few high validities as evidence of what the sophisticated clinical method can do in a study where it is given a chance to prove itself on grounds of its own choosing? I believe that we are. The psychologists who were our Judges I and II were considered to be good but not extraordinary clinicians, certainly no better than the best of the psychologists and psychiatrists who made the "naive clinical" predictions. They differed principally in that they had an adequate sample of data and had been through all the preliminary stages of studying the criterion in relation to the predictors in earlier groups of subjects. Moreover, they used systematic methods of analyzing the data, attempting to record all inferences and working with a set of intervening variables—personality constructs which were directly inferred from the test and interview data, and from which in turn the behavioral predictions were made. In a true sense, their clinical ratings were not naively based on unguided and uncontrolled judgment; they constituted a cross-validation of a whole structure of hypotheses about the determinants of psychiatric performance based on intensive prior study. Even so, our study left a great deal to be desired as a model of the sophisticated clinical approach—particularly on the scores of (*a*) a better job analysis, (*b*) a more broadly based, configurational approach to the design of manuals, and (*c*) a better stabilized criterion (see 3).

By way of contrast, a few more data before a final summing up. You remember that the battery of tests used in the first naive clinical predictive design consisted of the Wechsler-Bellevue, Rorschach, and Word Association tests. (The Szondi was also given, but was usually ignored; and the Strong Vocational Interest Blank was also routinely given, but was never scored in time to be used in the actual assessment for the Admissions Committee). The Rorschachs we gave were of course scored in the conventional way as well as by our special manual, and we thought it might be fun to see how some of the usual Rorschach scores would be related to the criterion. So we tried 14 scores and simple indices (like A%) with one class, and were surprised to find some rather high correlations. We decided therefore to see how a straight statistical-actuarial method of using the Rorschach would perform. Scrutinizing the table of intercorrelations between the Rorschach scores, we chose five of them that promised the greatest chances of success: DR%, number of good M, new F+%, F% and Stereotype% (scored after Rapaport, 13)—the last two because they looked as if they might be good "suppressor variables." A multiple regression equation was worked out to give the best linear combination of these variables to predict Over-all Competence; *R* for this Class was **.43** ($N = 64$). We noticed, however, that in the regression equation only the first two scores seemed to be playing any appreciable part, and in fact the multiple correlation using only per cent of rare details and number of good M was also **.43**. The other three were dropped from the formula, which was then tested on the Rorschach scores and criterion ratings of the first three classes. As expected, the correlation dropped out of sight on being cross-validated; with the new group of 116 subjects, it was .04.

The Strong Vocational Interest Blank, which (with an intelligence test) gave the best validities in the Kelly-Fiske study, likewise failed to yield any good predictor of competence in psychiatry. Even the special key ("Psychiatrist A") produced by Strong from a statistical analysis of thousands of blanks filled out by diplomates in psychiatry (15) failed to predict any of our criteria at a statistically significant level: no *r*'s as high as .2. This last finding deserves emphasis, because Strong's key was the product of a highly developed statistical technology, had an adequate numerical base, and had every opportunity to show what a pure actuarial method could achieve.

---

of **.36** or higher. (For a fuller discussion of the problem of contamination in these results, see [3].)

We made a further attempt to combine the best-predicting scores from the tests used in the standard battery into a regression equation. The R between Verbal IQ (Wechsler-Bellevue), Lawyer key (Strong VIB), DR% and No. Good M (Rorschach), and Overall Competence was .56 on the original group of 64; cross-validated on 100 cases, it dropped to .13.

SOME PRACTICAL IMPLICATIONS

If we had concentrated on an actuarial rather than a clinical approach and had come up with a simple, objective procedure that had a high and stable level of validity in predicting psychiatric performance, it could have been misused. It might have tempted many psychiatric training centers to adopt a single mold from which would have been cast a generation of psychiatrists, who would have had to meet the problems of the future with a standard set of resources derived from the past. The more successful we are in finding objective, impersonal, and statistical methods of selecting members of a profession in the image of its past leaders, the more rigid will be the pattern into which it is frozen.

For a concrete example, consider Strong's Psychiatrist A key again for a moment. It expresses the pattern of interests held in common by men who were diplomates in psychiatry at the end of the war, most of whom must have trained fifteen to twenty years ago. It should hardly be surprising that residents whose interests most closely approached this pattern tended to have skills as administrators and diagnosticians rather than as psychotherapists. If they had happened to achieve a high correlation with our over-all criterion, it might have helped populate American psychiatry of the 1960's and 1970's with near-replicas of the old state hospital superintendent.

It might be argued, however, that a similar result could have been expected if we had succeeded in providing explicit methods of clinically analyzing other types of data to select psychiatric residents. They too would have been based on a study of men who were successful at one time in history, and would have suffered the same danger of getting out of date. The answer is that even sophisticated clinical prediction never gets quite that rigid. Changes creep in; the result may be that validities gradually regress, or the drift may be determined by valid appraisals of newly important variables. Clinical methods are more flexible than their actuarial counterparts; they *can* be more readily modified by new studies based on observations of developing trends in the criterion. Moreover, valid clinical impressions can be obtained from an intensive study of a few known cases, while it takes large samples to set up or revise an actuarial system. There can be no guarantee that clinical methods *will* be kept up to date, of course, nor that the attempt to do so will not spoil their validities. Any predictive system needs constant overhaul and revalidation.

By sticking with the only capriciously accurate, sporadically reliable, and eminently flexible method of clinical judgment in selecting trainees, psychiatry will at least be able to keep in touch with developments in a growing and changing profession. Moreover, it will be able to maintain a healthy diversity within its ranks. There are many jobs to be done in psychiatry, requiring quite different kinds of men. There must be thoughtful men who like to sit in deep chairs and analyze patients all day long. There must be activists to organize new institutions and give inspirational leadership to groups of colleagues. Psychiatry needs many more men than it has whose main interest is in research and teaching, others to work with broad preventive programs in public health, group therapists, specialists in somatic treatments, and many more varieties of the general species. If the pure actuarial approach were to be seriously applied to psychiatry, it would be necessary to develop a formula for each of many different types of practice and to revise it constantly as new developments created needs for new types of practitioners. To do so would be impossibly expensive and laborious. Psychiatry is well-off, therefore, sticking with a basically clinical approach to assessment and prediction in selecting its members, but trying constantly to make it more scientific.

The important issue, however, is not what method of selecting its members is best for any particular profession, but the relative inertia of actuarial predictive systems and the maneuverability introduced when the generating of predictions is done by clinical judgment. This freedom is a source of weakness as well as strength; it enables the clinician to fall into errors of many kinds (to which statistical predictions are less subject) and also to adapt himself sensitively to all kinds of changing circumstances. When clinical methods are given a chance—when skilled clinicians use methods with which they are familiar, predicting a performance about which they know something—and especially when the clinician has a rich body of data and has made the fullest use of the systematic procedures developed by actuarial workers, including a prior study of the bearing of the predictive data on the criterion performance, then sophisticated clinical prediction can achieve quite respectable successes. I hope that clinicians will take some heart from our results, but I urge them to refine their procedures by learning as much as possible about statistical prediction and adapting it to their own ends.

To summarize: Meehl failed in his aim to mediate the statistical–clinical quarrel because he defined the issues in a way that perpetuates competition and controversy. The real issue is not to find the proper sphere of activity for clinical predictive methods and for statistical ones, conceived in ideal-type terms as antithetical. Rather, we should try to find the optimal combination of actuarially controlled methods and sensitive clinical judgment for any particular predictive enterprise. To do so, we must strike the right balance between freedom and constraint, a balance which may shift a good deal in one direction or the other depending on the nature of the behavior being predicted. But we can find such balances only if clinically and statistically oriented workers give up contentious, competitive attitudes and seek to learn from each other.

REFERENCES

1. BLENKNER, M. Predictive factors in the initial interview in family casework. *Soc. Serv. Rev.*, 1954, **28**, 65–73.
2. CRONBACH, L. J. Assessment of individual differences. In P. Farnsworth & Q. McNemar (Eds.), *Annual review of psychology*, Vol. VII. Stanford, Calif.: Annual Reviews, 1956. Pp. 173–196.
3. HOLT, R. R., & LUBORSKY, L. *Selecting psychiatric residents*. New York: Basic Books, in press.
4. HOLTZMAN, W. H., & SELLS, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, **49**, 485–490.
5. HORST, P. *et al. The prediction of personal adjustment*. New York: Soc. Sci. Res. Council Bull. 48, 1941.
6. KELLY, E. L., & FISKE, D. W The prediction of success in the VA training program in clinical psychology. *Amer. Psychologist*, 1950, **5**, 395–406.
7. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor, Mich.: Univer. of Michigan Press, 1951.
8. LUBORSKY, L. L., HOLT, R. R., & MORROW, W. R. Interim report of the research project on the selection of medical men for psychiatric training. *Bull. Menninger Clinic*, 1950, **14**, 92–101.
9. MCARTHUR, C. Clinical versus actuarial prediction. In *Proceedings, 1955 invitational conference on testing problems*. Princeton, N. J.: Educational Testing Service, 1956. Pp. 99–106.
10. MEEHL, P. E. *Clinical vs. statistical prediction*. Minneapolis: Univer. of Minnesota Press, 1954.
11. MEEHL, P. E. Clinical versus actuarial prediction. In *Proceedings, 1955 invitational conference on testing problems*. Princeton, N. J.: Educational Testing Service, 1956. Pp. 136–141.
12. OSS Assessment Staff. *Assessment of men*. New York: Rinehart, 1948.
13. RAPAPORT, D , GILL, M. M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. II. Chicago: Yearbook Publishers, 1946.
14. SARBIN, T. R. A contribution to the study of actuarial and statistical methods of prediction. *Amer. J. Sociol.*, 1943, **48**, 593–603.
15. STRONG, E. K., JR., & TUCKER, A. J. The use of vocational interest scores in planning a medical career. *Psychol. Monogr.*, 1952, **66** (9), No. 341.
16. WITTMAN, M. P. A scale for measuring prognosis in schizophrenic patients. *Elgin Papers*, 1941, **4**, 20–33.

Received April 15, 1957.

SOCIAL INFLUENCE PHENOMENA IN A TASK-ORIENTED GROUP SITUATION[1]

J. S. KIDD[2]
Northwestern University

The roughly similar results for social influence on psychophysical judgments (Sherif [9]) and on attitudes and values (Crutchfield [3]) seem to justify the use of the simpler and more easily quantified psychophysical judgments to bring important social phenomena into the laboratory for controlled study. The present experiment was designed to explore the effects of several situational variables on the phenomena of social influence in the behavioral context of a task-oriented group. It was assumed for the purposes of this study that the mechanisms involved in social influence would be manifest subsequent to the exposure of the subject (*S*) to group participation and that the behavior to be influenced need not be relevant to the group task. Social influence is defined here as a change in a response following the presentation to *S* of a response standard of apparent social derivation. A change leading to a higher degree of conformity between the response and the standard indicates positive social influence; a change in response in the direction of less conformity is taken as negative social influence.

The three variables tested were the size of the group, the duration of group participation, and the identifiability of the source of the response standard. Group size was selected as a variable in part because of the apparent incompatibility between the findings of Asch (1) and those of Goldberg (4). Asch found an increase in social influence as the size of the (reporting) group increased. Goldberg, using a different experimental approach, was not able to demonstrate this effect. Interest in duration of participation stemmed from empirical work on voluntary social groups by Bovard (2), who found that conformity (social influence) behavior could not be predicted from length of membership. The present study sought to extend Bovard's findings by considering the same variable against the background parameters provided by nonvoluntary groups. Identifiability of the source of the response standard was suggested as an important variable by an interpretation of some of the general conclusions of Hovland, Janis, and Kelly (5) concerning the effect of personal presence in persuasion.

The three experimental hypotheses tested were that positive social influence increases (*a*) as group size increases, (*b*) as duration of participation increases, and (*c*) if the source of the response standard is identifiable rather than remaining relatively anonymous. These hypotheses, while not entirely consistent with the empirical findings previously cited, provide an internally unified approach compatible with much contemporary theory.

As a secondary consideration in this study, task performance measures were obtained from the various sized groups in order to test the hypothesis that task effectiveness increases with group size.

METHOD

Design and Subjects

The variables of group size and duration of participation were each evaluated at three levels. Groups of two, four, and six members were compared. The durations of group participation selected were 1, 2, and 3 hrs. The variables were incorporated in a 3 x 3 factorial design. In order to test the hypothesis regarding the identifiability of the source of the response standard, an additional (nonfactorial) condition was introduced that involved the observation of solitary *S*s.

In the conditions of two- and' six-member groups there were 18 *S*s in each cell; in those for four-member groups, 20 *S*s. Fourteen *S*s were employed in nonfactorial condition. In all cases, individuals rather than groups were considered as the responding entities. The total sample *N* of 182 was composed of 100 female and 82 male undergraduate students from beginning social science classes at Northwestern University. Sampling bias was minimized by randomizing the sequence of conditions and by assigning *S*s to conditions in the order in which they volunteered for participation in the experiment.

---

[1] This study is based upon a dissertation submitted to the graduate school of Northwestern University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is grateful to Drs. Donald J. Lewis and Donald T. Campbell for their patient assistance and sympathetic sponsorship.

[2] Now with the Laboratory of Aviation Psychology, Ohio State University.

TABLE 1

NUMBER OF CASES AND MEANS FOR MEASUREMENT OF SOCIAL INFLUENCE

Group Size	Duration of Contact			Mean[a]
	1 Hr.	2 Hrs.	3 Hrs.	
One member		14		4.00
Two members	18	18	18	5.18
Four members	20	20	20	5.27
Six members	18	18	18	5.54
Mean	5.02	5.63	5.34	

[a] The scale used was from 1.00 through 9.00 with the higher score indicating greater positive social influence.

TABLE 2

ANALYSIS OF VARIANCE OF SOCIAL INFLUENCE SCORES

Source of Variance	*df*	Mean Square	*F*
Group size	2	2.00	.51
Duration of contact	2	5.03	1.28
Interaction	4	6.50	1.65
Within groups error term	159	3.94	

Procedure

All *S*s were first informed that the purpose of the experiment was the determination of work team effectiveness in accomplishing simple tasks. They were told that on completion of the experiment they would be asked to participate in an additional study, but the nature of the second study was not revealed.

The experimental sessions that followed were of 1-hr. duration. Those groups meeting for two or three sessions came together at intervals of 48 hrs. During each session the groups (or individuals) worked on three tasks: an anagram task, a jumbled sentence task, and a jigsaw puzzle task. In the group situation, the task was structured so as to force cooperative activity, by distributing task materials equally through the group and requiring that the group product contain contributions from each group member. In the condition employing single *S*s, all task materials were presented simultaneously. Knowledge of task results was held at an artificial constant value by *E* in order to minimize variability for any "group success" effect (6).

Following the final task session, *S*s were informed that the main experiment was complete. They were asked then to help out in a brief study of a simple psychophysical nature. The *S*s were seated along the back of the experimental room equidistant from the stimulus source and separated from one another by a minimum of seven feet. The room was partially darkened. The *S*s were provided with response slips and were asked to observe a flickering light for five seconds. Their required response was to estimate the number of flickers in that period. When their estimates were complete, *E* collected the response slips and after a short interval returned to each *S* a fresh response slip upon which was written a purported "group average estimate." The *E* called attention to this figure, indicated that it was the average estimate of the group and that it was presented to give "some idea of what other members of your group had judged the correct total to be." The *S*s were then required to observe the flicker again for five seconds and note a second estimate.

In each case, the purported group average estimate was fictitious. A standard factor of .5 or 1.5 (depending on whether the *S*'s original estimate had been higher or lower than the true value) was applied and the result was reported to each *S* individually as the group figure. This procedure was followed in order that the degree of deviation from the response standard would be proportionately the same for each *S*. The situation was structured so that the discrepancies between each *S*'s "group average estimate" score could not be noted.

For *S*s who worked through two sessions of the task phase alone, the instructions at the terminal measurement of social influence were intended to provide a nonidentifiable source for the response standard. Following *S*'s first flicker estimate, he was given a second slip upon which a figure labeled "average estimate" appeared, and informed that this estimate represented his own original estimate averaged in with the estimates of "some other people in this situation." Neither the number nor any other characteristic of the fictitious others was indicated, thus providing a relatively anonymous or nonidentifiable response source.

For all *S*s, a brief description of the true nature of the study was given following the completion of the social influence measurements.

RESULTS

Social Influence

The effects of the two main variables of group size and duration of participation upon the social influence response were evaluated by both analysis of variance and by the nonparametric χ^2 contingency test. The results of both methods of analysis fail to reveal reliable differences attributable to the experimental conditions. A summary of the data and the statistical analyses is presented in Tables 1, 2, and 3.

The original raw data were collected in the form of percentage change in response with algebraic signs to indicate direction of change. The complete raw score distribution was skewed sharply toward the positive side. For purposes of the analysis of variance, a log transformation was applied and the resultant transformed scores were grouped in nine class intervals. The resulting scale of social influence was found to be suitable for analysis.

Tests of the main effects and interactions using analysis of variance were negative. The small trend in favor of greater social influence with increasing group size was not significant.

The results of the analysis of variance were confirmed by the use of χ^2 contingency tests on the raw data, which fell into a natural trichotomy: positive social influence, no change, and negative social influence. A 3 x 3 contingency table arrangement was used to test the effects of the two continuous variables. The variable of group size yielded a χ^2 of 4.75 which is not significant. The duration of group participation variable gave a χ^2 of 1.95 which is also not significant.

In order to test the hypothesis regarding identifiability of the source of the response standard, a comparison was made between the influence scores of *S*s who participated as individuals and *S*s in the two-member group condition having the same duration of participation (in this case two task sessions). The difference between means was evaluated by *t* test. The *t* ratio obtained was 4.01 which for 30 *df* is significant beyond the .01 level. Further confirmation was accomplished by pooling the response data for the three conditions of group size (two, four, and six members), and comparing the resultant mean score with the mean score from the condition of low identifiability of the source of the response standard. The *t* ratio obtained in this comparison was 5.14 with 68 *df*, which was again significant beyond the .01 level of probability. In both instances, the difference between the means was in the direction of greater positive social influence when the source of the response standard was identifiable. A summary of *t*-test comparisons is presented in Table 4.

TABLE 3
FREQUENCY ANALYSIS OF SOCIAL INFLUENCE RESULTS

Direction of Social Influence	Group Size 2	Group Size 4	Group Size 6	Totals
+	39	41	43	123
0	7	12	9	28
—	8	7	2	17
Total	54	60	54	168
	$\chi^2 = 4.75$, $df = 4$			
	Duration of Participation			
	1 Hr.	2 Hrs.	3 Hrs.	
+	39	43	41	123
0	9	9	10	28
—	8	4	5	17
Total	56	56	56	168
	$\chi^2 = 1.95$, $df = 4$			

TABLE 4
COMPARISON OF MEAN SOCIAL INFLUENCE SCORES BY *t* TEST: IDENTIFIABLE VERSUS NONIDENTIFIABLE SOURCE OF RESPONSE STANDARD

Condition	Mean Social Influence Score	σM	*t*	*df*
Nonidentifiable source	4.00	.237		
Identifiable source (Two-member group)	6.00	.434	4.01*	30
Identifiable source	5.63	.254	5.14*	68

* $P < .01$.

Task Efficiency

To test the effect of group size on task effectiveness, performance measures were obtained on two of the three tasks assigned to the groups and individuals across conditions. The anagram task required differential instructions, so it was eliminated from consideration. The remaining tasks, sentence completion and jigsaw puzzle, were analyzed in two ways: with the whole group considered as the producing unit and with the individual member considered as the producing unit.

TABLE 5
ANALYSIS OF VARIANCE OF TASK EFFECTIVENESS OF VARIOUS SIZED GROUPS ON SCRAMBLED SENTENCE JIGSAW PUZZLE TASKS

Task	Source of Variance	*df*	Mean Square	*F*
Scrambled Sentence	Team Size	3	25.6	57.78*
	Within conditions error term	61	.443	
Jigsaw Puzzle	Team Size	3	661.0	21.05*
	Within conditions error term	61	31.4	

* Significant beyond the .01 level of confidence.

For the sentence completion task, efficiency was measured by sentences completed per unit time. No significant differences attributable to group size were observed when analysis was based on the group as the producing unit. Considering the individual as the producing unit, a significant *F* ratio of 57.78 was obtained ($p < .01$). For the jigsaw puzzle task, completion time for the three puzzle solutions accomplished in a single session was pooled to give a mean completion score. As in the sentence completion task, when the data were analyzed with the individual taken as the producing unit, the F ratio was 21.05, which is also significant beyond the .01 level. In both instances, the trend was one of decreasing task effectiveness as the number of co-acting group members increased. Table 5 is a summary of the analyses of variance performed using the individual member as the producing unit.

DISCUSSION

Social Influence

Contrary to the expectations expressed in the original hypothesis, group size was not shown to be effective over the range of two, four, and six members. This result confirms and extends the findings obtained by Goldberg (4) for two- and four-member groups. It is inconsistent, however, with Asch's results (1). All three studies nevertheless show the same general trend. Goldberg's *S*s in four-member groups were influenced more than their counterparts in two-member groups—but not significantly more. The finding of a consistent, though as yet nonsignificant, trend such as this should encourage investigators to continue attempts to isolate the effect of group size upon social influence phenomena. It is certainly possible that the lack of clearly positive results reflects the lack of precision in measuring instruments.

Again, contrary to the original hypothesis, duration of group participation was not revealed as an effective factor in social influence. This result confirms Bovard's finding (2) in a substantially disparate set of circumstances. The fact that in both studies all groups were relatively homogeneous with respect to age, college attendance, etc., is probably a relevant consideration. A conclusion that overreaches the data but seems reasonable is that among *S*s with well-established habits of peer-group interaction, the social influence response reaches an asymptotic value within a few moments of group contact.

With regard to the comparison of the easily identified as opposed to the relatively nonidentifiable source of the judgmental standard, there were measurably greater social influence effects when the source of the standard was known to be fellow group member(s). This is not to say that conformity did not occur when the source of the standard was anonymous. On the contrary, there was still a measurable tendency, in some *S*s at least, to modify their judgments in the conforming direction even when they have no ready referent for the source of the standard. But when this standard arises within a group of which they are even temporarily and superficially a part, the conformity response is considerably strengthened.

Task Efficiency

The data fail to confirm the hypothesis that efficiency would increase as size increased when the group was taken as the producing unit. This finding is inconsistent with the results obtained by Shaw (8) and by Taylor and Faust (10), but agrees with those of Moore and Anderson (7).

The contrary finding for two of the tasks that, with the individual taken as a producing unit, efficiency declines as group size increases tends to confirm the observations of both Taylor and Faust and of Moore and Anderson when they treated their data in individual rather than group terms. The finding appears to be rather consistent over a variety of tasks and under quite different experimental conditions.

Taking the group as the unit of production, however, leads to a marked disagreement. The Shaw and the Taylor and Faust studies found improvement with increased group size; in the Moore and Anderson and the present investigations, no differences were found between individuals and groups of various sizes. It is quite possible that the discrepancy is attributable to factors in the tasks or the general experimental operations. But conjectures can be advanced that resolve the discrepancy on theoretical grounds.

If one assumes that the group members can exert a restraining effect or a releasing effect on their fellow group members, the possibility exists that when the problem-solving activity is sustained over a considerable duration (as it was, for instance, in the Shaw study) a disproportionate effect is produced by the members with the greatest ability to solve the particular problem at hand. If the experiment is started with two samples equivalent in problem-solving ability, and one sample is divided into groups, the group members who are poor in problem-solving ability should affect the final outcome to a lesser degree than should persons of low ability who work on the problems alone. Where groups have been found superior to individuals in task effectiveness, it is suggested that the outcome may be attributable to the fact that in the group situation some members were restrained from acting and others were encouraged to act—depending upon their apparent task ability.

SUMMARY

The effects of group size, duration of group participation, and identifiability of the source of a response standard were evaluated with respect to a measure of social influence. A total of 182 *S*s were observed. A factorial design was used to test group size of two, four, and six members with duration of participation at 1, 2, and 3 hrs. An additional condition with single *S*s was introduced to evaluate the effects of the identifiability of the source of the response standard. Cooperative group activity on three tasks was followed by a single measurement of social influence on a nontask-relevant psychophysical judgment.

For the particular situation studied, neither group size nor duration of group participation had significant effects on social influence. However, identifiability of the source of the response standard was accompanied by increased social influence.

A secondary area of investigation concerned the role of group size on task effectiveness. The results from a sentence completion and a jigsaw puzzle task were evaluated in this connection. With the group considered as the producing unit, increasing group size did not facilitate task performance. On the contrary, with the individual considered as the producing unit, increasing the number of co-acting members actually decreased the efficiency of task performance.

REFERENCES

1. ASCH, S. E. Some forms of interpersonal influence. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (Rev.) New York: Holt, 1952. Pp. 2–11.
2. BOVARD, E. W. Conformity to social norms and attraction to the group. *Science*, 1953, **118**, 598–599.
3. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
4. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.
5. HOVLAND, C. E., JANIS, I. L., & KELLY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.
6. KIDD, J. S., & CAMPBELL, D. P. Conformity to groups as a function of group success. *J. abnorm. soc. Psychol.*, 1955, **51**, 390–393.
7. MOORE, O. K., & ANDERSON, S. B. Search behavior in individual and group problem solving. *Amer. soc. Rev.*, 1954, **19**, 702–714.
8. SHAW, M. E. A comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.*, 1932, **XLIV**, 491–504.
9. SHERIF, M. *The psychology of social norms.* New York: Harper, 1936.
10. TAYLOR, D. W., & FAUST, W. L. Twenty questions: efficiency in problem solving as a function of size of the group. *J. exp. Psychol.*, 1952, **44**, 360–368.

Received December 3, 1956.

PERSONALITY AND PSYCHOPATHOLOGY[1]

E. I. BURDOCK, SAMUEL SUTTON, AND JOSEPH ZUBIN
Biometrics Research, New York State Department of Mental Hygiene and Columbia University, New York

THE hypothesis that the phenomena of mental disease are the outward manifestations of impaired brain function suggests a systematic exploration of the various dimensions of behavior to reveal the extent and the depth of defect. It rests with biometrics, the science which applies quantitative analysis to the structure and function of living organisms, to provide a map for such a survey by pointing out the main avenues along which progress may be expected, to integrate observations obtained at different levels of function, and to devise new tools for the extraction of significant information.

BIOMETRIC MODEL FOR ANALYSIS OF BEHAVIOR

An analysis of individual behavior can be organized by allocating observations among five major levels of activity: (*a*) physiological, (*b*) sensory, (*c*) perceptual, (*d*) psychomotor, and (*e*) conceptual.

Observations of such autonomic functions as heart rate, blood pressure, rate of respiration, production of sweat or saliva, etc., tap an area of behavior which can be measured without the overt cooperation of the subject (*S*) and even without his attention. In particular, while such measures are noticeably sensitive to external stimulation, they can nevertheless be obtained when the *S* is at rest, in what might be called the "idling" state, so that it would be possible to establish base lines of psychophysiological activity against which the influence of extrinsic stimuli could be more accurately assayed than at present.

Sensory activity refers to the modality-specific component of the complex of behavior involved in the functioning of a receptor organ. The immediate experience correlate of sensory activity in the human sensorium is a sensation, e.g., the mere awareness of light when the retina is illuminated. The neural conditions for a sensation are as follows:

1. Electromagnetic or gravitational energy impinging on a sense organ or free nerve ending.
2. Conduction of a nerve impulse by the specific afferent system from the receptor.
3. Collateral nonspecific afferent conduction in the reticular activating system (3, 31).
4. Efferent feedback regulation of the sensitivity level of the receptor (9, 11, 22).
5. Integration of peripheral, central, and autonomic neural activity at the level of the diencephalon (26).

The absence of the first condition, external energy, distinguishes a hallucination from a true sensation. One test that has been suggested for this distinction is the occurrence of after-images (41, pp. 6–9); but there is evidence that both eidetic images (17) and visual hallucinations (40) may be followed by after-images. Pure sensation is probably infrequent. Seeing light, feeling pain, hearing sound, etc., with no further qualification approximate the experience, but probably involve some conceptual abstraction on the part of the *S*.

In contrast to sensation, which is concerned with the unique quale of a sense modality, perception is an activity of the organism which superposes on the sensory function, scanning, focusing, and following mechanisms that are experienced as discrimination and that are oriented to reaction. In addition to the neural events involved in sensation, perceptual behavior incorporates primary efferent pathways into the circuit. Perception is thus the experience correlate of the integration of the primary afferent and efferent systems (30).

When overt movements accompany the perceptual act, the behavior is psychomotor. It is likely that some psychomotor activity accompanies most perceptual behavior.

The expression, conceptual behavior, is best reserved for those obscure components of the cortical network which mediate linguistic and metalinguistic activity.

Elsewhere (4) it has been pointed out that these five categories comprise an ascending hierarchy with physiological behavior providing the vegetative substrate and with conceptual behavior involving at least the first three

---

[1] Presented in part before Section I of the AAAS, New York, December, 1956. The preparation of this paper was facilitated in part by research grant, M-586 (C3) Psychological Prognosis in early and chronic schizophrenic patients, from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

levels as precursors or concomitants. A comprehensive account of an individual's mental functioning ought to include representative samples from each distinguishable level of behavior.

It is just as inadequate to evaluate psychopathology solely in terms of conceptual functioning as it would be in physical science to account for the phenomena of solar radiation only in terms of terrestrial combustion and without regard to the behavior of nuclear plasmas at extreme elevations of temperature. Indeed, the enormous energy expended by the sun without appreciable loss of mass remained a mystery until the submolecular aspects of the system were brought into meaningful relation with the molar observations by Bethe's hypothesis of the carbon cycle.

For psychopathology, biometrics has the responsibility of searching through the whole range of human behavior for the critical relations between different levels of function which will throw light on the mystery of mental disease. It is already abundantly clear from the history of psychopathology that those investigators who confine their attention to a single level of behavior cannot hope to account for the problem in all its complexity.

A map of the world of psychopathology needs meridians to distinguish the experimental techniques that can be applied across the latitudes of behavior. In studying the behavior of an individual, the experimenter may adopt one of three methods: First, he may merely observe the *S* without intruding in any way into the natural flow of behavior; second, he may seek to engage the *S* in some specific activity; or third, he may modify or distort the intrinsic functioning of the *S* and then proceed to examine the residue of behavior by one of the above two methods. The first procedure, naturalistic observation, is propaedeutic to experimentation. Observation and description provide a context from which hypotheses can emerge for testing under controlled conditions. Even the ethologists (27), who rightly insist on the importance of studying animal behavior in naturalistic settings, nevertheless do not hesitate to manipulate the environment so as to determine the limits within which a particular act, such as a duck's retrieving an egg, will be evoked.

The second method, the controlled experiment, has been the traditional method of choice of the experimental psychologist. A significant feature of experimental procedure is that the *S* is removed from his natural habitat to the artificial environment of the experimental laboratory.

In order to explore the intrinsic behavioral repertoire of an individual, the experimenter usually seeks to bring the *S* into reaction with some limited aspect of the environment which can be varied qualitatively or quantitatively. He then observes the activity of that segment of the organism which is peculiarly sensitive to the slice of the environmental surroundings that the experimenter is manipulating. The experimenter frequently establishes a baseline for his study by initiating his record before he has introduced the variable quality or quantity. The behavior of the organism corresponding to such a zero point may appropriately be described as in an "idling" state. With the organism in the idling state, the experimenter may conduct observations at any one of the five behavioral levels described above (physiological, sensory, perceptual, psychomotor, and conceptual.) Characteristically, the experimenter restricts his observations to one of these levels. Indeed, in this age of specialization, an entire professional lifetime may be devoted to the detailed study of a single behavioral level, or, more commonly, of some particular subdivision of such a level, e.g., auditory sensitivity (sensory), visual form discrimination (perceptual), word associations (conceptual), etc.

Just as the domain of individual behavior can be analyzed into five levels, so too can the parameters varied by the experimenter be differentiated into two broad classes above the zero point with its associated "idling state": energy variables and signal variables.

Energy variables are stimuli which furnish the necessary driving energy for a response. Characteristically, the stimulus sequence is graduated in units of intensity, duration or wave length (or its reciprocal, frequency). As the response is powered by the input energy, the experimenter seeks to express the corresponding activity in the organism as a mathematical function of energy. Energy variables are commonly used to study sensory functions, such as thresholds or j.n.d.'s. The energy may be appropriate, i.e., within the range to which the sense organ is particularly responsive, as light to the eye, sound to the ear; or it may be *inappropriate*, i.e., outside the energy range for which the organ is specially adapted, as electrical stimulation above the retina to produce phosphenes, or caloric stimulation of the vestibular apparatus.

Signal variables, on the other hand, provide a relatively insignificant fraction of the energy released in the response. They are triggering devices which vary in spatial or temporal organization. Correspondingly, the response, deriving its motive power from energy intrinsic to the organism, can no longer be represented as a function of input energy.

The simplest kind of signal variable is the *configuration* in which the stimuli consist of spatio-temporal patterns or arrangements. Variations of juxtaposition or of time interval produce corresponding changes in behavior, as in the visual discrimination of shape or the auditory discrimination of timbre.

A more complex kind of signal variable is the *sign*. A sign is any event which implies some other event. More generally, a sign places a probability valuation on the event it designates. Thus, smoke is a sign of fire; a footprint in the sand signifies that someone has stepped there; a certain residue of C^{14} in a fossil plant is a sign of its age. Natural signs are either precursors or consequences of the events they designate. It is, however, a universal characteristic of human behavior to employ sets of arbitrary conventional signs for cultural purposes. For example, an arrow indicates direction; a red light warns of danger; a warbling tone on a siren gives the all clear. A set of signs may be said to comprise a code when arranged in some lexical order over against a corresponding set of designated events. The process of transmitting and receiving such encoded signs is the signal engineer's definition of communication. In the psychological laboratory, if the experimenter places an arbitrary sign (CS) in temporal or spatial contiguity with a reinforcing event (US), the sign by its presence or absence establishes a pattern, or configuration, which may be varied according to a predetermined sequence. Classical Pavlovian conditioning, animal experiments in visual or tactile discrimination, and delayed reward experiments employ this technique.

A *symbol* is a sign which designates an aggregate or class of events. Symbols are by their very definition abstract, removed from direct reference to particular events. Language, i.e., the systematic designation of classes of objects, attributes, or events by accidental patterns of acoustic energy, is in essence a complex communication code composed of symbols. In all likelihood, linguistic behavior is the prototype from which symbolism in art and music is derived. In origin, all music is "program music," i.e., has content reference, and all art is representative. The complex experience evoked by a work of art through the interplay of elements of pure design with subtly suggestive allusions to persons and events is probably prepotent in such devices as the Rorschach and the TAT. These techniques should therefore be viewed as presentations of symbol variables, and the corresponding behavior should be designated conceptual rather than perceptual. Other instances of behavior evoked by symbols are found in the interview and in the questionnaire. Table 1 shows examples of measurable activities as functions of stimulus variables.

The Modified Organism

On the premise that the whole is the sum of its parts, physiologists have long had the practice of studying isolated systems, e.g., the nerve-muscle preparation, the perfused liver, etc. Taking a leaf from this common physiological procedure, experimental psychologists and psychiatrists have in recent years begun to study the organism's behavior after first introducing certain modifications or distortions of intrinsic functioning. This has been done in three ways: imposition of a load, intervention in the internal neurohumoral environment, and ablation or destruction of part of the organism.

1. *Loads.* A load is defined as an externally imposed force which induces a distortion or strain and a counteracting stress. In a physical object, e.g., a wire with a weight attached, the restorative stress is proportional to the distortional strain, in accordance with Hooke's Law. Similarly, when a psychological load is imposed on an organism, the organism too undergoes a strain which is countered by restorative stress, although it is still moot whether an analogue of Hooke's Law can be established for psychology. Examples of psychological

1-9-70
J 279

TABLE 1

EXAMPLES OF MEASURABLE ACTIVITIES AS FUNCTIONS OF STIMULUS VARIABLES

Level of observed behavior	Stimulus Variables													
	Idling State		Energy Variables				Signal Variables							
	None		Appropriate Energy		Inappropriate Energy		Configurations		Signs		Symbols			
	Variable	Function	Variable	Function	Variable	Function	Variable	Function	Variable	Function	Variable	Function		
Conceptual	—	Reverie and fantasy	Uniformly diffused light	Fantasy	Inversion of gravitational attraction	Subjective account of experience	Aircraft forms or silhouettes	Recognition of identity of forms	Practical trouble-shooting test	Diagnosis of trouble	Stimulus word in association test	Association to stimulus word		
Psychomotor	—	Spontaneous movement	Painful stimulus	Arm withdrawn	Electroshock	Movements of limbs	Star-shaped maze	Mirror tracing	Classical delayed response stimuli in animal experimentation	Successful response by animal subject	Reinforcement of affect in focused interview	Electromyographic response		
Perceptual	—	Spatial and temporal orientation	White noise	Orientation to direction of sound	Pressure stimulation above retina	Phosphene	Rotating Benham Disc	Subjective color experience	Usual visual alternatives in animal discrimination experiment	Selective response of animal subject	Musical chords	Pitch discrimination		
Sensory	—	Background noises; cortical gray	Light of graded intensity	Threshold response	Electrical stimulation of thermal receptors	Warmth or cold sensation	Patterned light stimuli	Visual threshold	Infant's faint cry	Mother's auditory threshold	Words or sentences presented tachistoscopically	Visual threshold		
Physiological	—	BMR; basal EEG; basal PGR	Increase in carbon dioxide concentration	Change in rate of respiration	Pressure on carotid sinus	Change in heart rate	Photic driving	Change in EEG pattern	Bell ringing in Pavlovian conditioning	Salivation	Verbal instructions to prevaricate	Effect on PGR		

loads are threat, with consequent strain of anger or fear (5); tissue damage, with the strain reflected in the inflammatory process; deprivation of air, food, water, sensory input, etc. (2, 15); and intense physical effort or prolonged exposure to extremes of temperature. Just as an excessively severe physical load will produce a permanent distortion in a wire, so too can a psychological load strain an organism beyond its elastic limit, so that subsequent strains must be measured from a new baseline. The consequences of psychological loads which are excessively severe or unduly prolonged are seen, for example, in "combat fatigue," in the General Adaptation Syndrome (39), or in the adrenal hypertrophy which follows intense uninterrupted physical exertion (35).

2. *Neurohumoral intervention.* The application of drugs or brief electrical current to produce activity in the organism has become increasingly popular with experimenters. Drugs like mescaline (18), adrenochrome (14), and LSD (42) have been used to evoke bizarre behavior or to initiate anomalies of subjective experience in the hope of uncovering a biochemical key to psychopathology. Despite many suggestive analogies, however, their effects are obscured by the wealth of interactions which they set up in the organism. The difficulties to be surmounted before a biochemical understanding of psychopathology can be expected may be appreciated when it is realized that the biochemistry of so common an agent as aspirin is still obscure. The initiation of behavior by intracranial electrical stimulation, first used for mapping the brain (8), has recently become a method of choice for the production of activity or affect (6, 32). But here again the relevance of these new techniques for the behavior of the intact organism remains to be demonstrated.

3. *Ablation.* The method of ablation, by which a part of the brain substance or of the neurohumoral network is excised, has made many substantial contributions to the understanding of the structure and function of the nervous system (1, 19, 38). However, the caution cannot be repeated too often that what is subsequently observed is the residual behavior of all the rest of the animal, including whatever compensatory or suppressed mechanisms may have been released. Inferences as to the role of the excised part in behavior remain at best dubious. This is particularly true because of the difficulty of assessing the side effects of trauma, degenerative changes, bleeding, anesthesia, etc.

Of the three methods for modifying the organism, the imposition of a load seems at present the most promising for psychopathology because it usually brings about a response that differentiates between mental patients and normals. Neurotics and early psychotics tend to react with excessive strain, while chronic schizophrenics show a deficiency in reactivity. In studying behavior under load, some care must be exercised to avoid confusing the load with the independent variable. A load which causes tissue damage characteristically takes the form of an energy variable. But a psychological threat may present itself as a sign or a symbol. Particularly when one is studying behavior of the "idling" state under load, there is a tendency, because of the absence of an independent variable, to think of the load as the variable. However, if the term *load* is restricted to those environmental factors which menace the integrity or existence of the organism, the distinction will serve a useful purpose.

INTEGRATION OF OBSERVATIONS

It has been noted for some time that the usual performance of the mental patient may not differ from the performance of normals, provided attention and motivation are sustained. It is quite possible that in chronic schizophrenics the prolonged strain induced by the load which the disease itself constitutes has brought about a permanent distortion in the patient's adaptive capacity so that he displays reduced sensitivity under any additional load. It is, therefore, important that autonomic base lines be established for patients vis-à-vis normals under the idling state and that other variables be studied both from the point of view of usual functioning and from the point of view of functioning under a load. The hypersensitivity displayed by early schizophrenics under load in contrast to the apathy of the chronics has already been noted. It may very well be that no single aspect of behavior will alone distinguish the patient from the normal, but that the pattern of function will be distinctive (23). An additional point that has been noted (47) is that patients whose conceptual functioning is depressed below their perceptual function are more likely to improve than those with conceptual behavior elevated above perceptual.

TOOLS AND TECHNIQUES OF RESEARCH

A program designed for the intensive exploitation of these new leads in psychopathology is represented in Table 2 in which certain techniques that have shown promise in preliminary undertakings are classified according to the schema previously described. The omission of techniques for the study of sensory activity is justified by the fact that sensory thresholds have not been found to bear any relation to outcome of mental illness. On the other hand, it should be emphasized that techniques which have failed to discriminate patients from normals under ordinary conditions may yield significant differences when the subjects are tested under a load, such as psychological threat, sensory deprivation, or extreme of temperature. The imposition of a load reveals the lowered tolerance and decreased resilience of the patient.

Physiological Techniques

The increasing evidence that schizophrenics may differ from normals in essential physiological function makes it urgent that experimental psychopathology give greater attention to the physiological level of behavior. The following two techniques represent efforts to measure autonomic components of physiological activity.

1. *Pupillography*. The method of recording pupillary reflexes developed by Lowenstein (28) presents an unusual opportunity to study the interaction between the sympathetic and parasympathetic divisions of the autonomic system, since presumably both divisions modulate the pupillary mechanisms. In this technique, serial photographs of the eyes are made under infrared illumination after dark adaptation and during controlled exposures to light. The diameter of the pupil is measured from the film and plotted against time. Lowenstein and his co-workers have performed numerous animal experiments and have assembled a large body of clinical data which point to relationships between pupillographic curves and specific lesions or dysfunctions of the central nervous system. This technique was applied to the study of schizophrenia in a pilot investigation by Lowenstein and Westphal in 1933 before the method had been satisfactorily standardized. The curves they obtained for catatonic schizophrenics were strikingly different from those of normals. The diminution of the pupillary reflex on repeated evocation is more pronounced and appears earlier in patients. Moreover, in normals, but not in schizophrenic patients, the initial level of response can be restored by inducing startle, i.e., when a sudden intense sound is introduced as a load, patients react differently from normals. It has also been noted that the pupillograms of schizophrenic patients show striking similarities to those of patients with extrapyramidal lesions.

2. *Pattern of autonomic responsivity*. Many investigators have reported differences between schizophrenics and normals in autonomic responsivity, particularly under conditions of stress-producing load. There is some indication that acute cases tend to overreact, chronic cases to underreact. However, the equivocal results of a substantial number of studies point up the need for further investigation. It is clear that the degree of activation of autonomic mechanisms is correlated with the strength of the stimulus, or load, ranging from simple alerting responses all the way to the severe and prolonged reactions evoked by combat fatigue. Research by Lacey (23) and by others has also shown that an individual's pattern of autonomic responses may be invariant despite great differences among individuals.

Taken together, these findings suggest the need to sample patterns of autonomic responsivity at several points along the intensity dimension. It may be that failure to find differences between schizophrenics and normals on a single response variable is caused by individuals with contrasting autonomic patterns who cancel out one another's responses in both groups. A polygraphic study of schizophrenics and normals provides a technique for determining the autonomic pattern as a basis of comparison. Such a study should include techniques from each class of stimulus variables, beginning with the idling state as a baseline and progressing through the energy operators and the signal operators. In particular, it is to be observed that the imposition of the load constituted by sensory deprivation may be expected to magnify the differences to be found between patients and normals.

TABLE 2

Classification of Techniques for Study of Early and Chronic Schizophrenics

Technique	Variable		Function		Load	Reaction
	Class	Description	Class	Description		
Pupillography adaptation and restitution	appropriate	alternating light and dark	autonomic	contraction and dilation	sudden intense sound	restitution
Pattern of autonomic responsivity	none (idling state) appropriate configural sign symbol	— flash of light to eye Archimedes Spiral colored light as CS interviewer's activity in affect-focused interview	autonomic	heart rate, blood pressure, P.G.R., respiration rate, etc.	sensory deprivation	corresponding changes in functions
Weight judgment unanchored anchored	configural configural	weight weight	perceptual perceptual	discrimination of wts. discrimination of wts.	— —	
Flicker-fusion	configural	intermittent light	perceptual	discrimination of "steady" from "flicker"	—	
Necker cube	configural	ambiguous figure	perceptual	discrimination of alternative figures	—	
Archimedes spiral rotation end of rotation	configural configural	visual pattern visual pattern	perceptual perceptual	perception of movement reversal of movement	— —	
Digital stimulator	appropriate	tactual stimulation of more than one finger	perceptual	discrimination between stimulated fingers	cold pressor	change in discriminative accuracy
Color-form test	configural	visual pattern	perceptual	report of direction of apparent movement	—	
Purdue pegboard	configural	pattern of pegs and holes	psychomotor	placement of peg	—	
Tapping reaction time tapping	sign configural	colored light pattern of tapping targets	psychomotor psychomotor	movement speed of motion	— —	
Reaction time shift	sign	visual and auditory	psychomotor	movement	threat	difference in latency
Delayed auditory feedback in reading	symbol	previous speech segment	psychomotor	articulation movements	pin prick	change in rate of articulation

Metalog	symbol	word	conceptual	definition	—	
Analogies	symbol	word	conceptual	eduction of correct analogue	—	
Essential differences	symbol	word	conceptual	discrimination of "different" word	—	
Porteus Maze	configural	visual pattern	conceptual	choosing direction	—	
Stroop I	symbol	word	conceptual	reading	shock	difference in latency
II	sign	color	conceptual	naming		
III	symbol	color word printed in conflicting color	conceptual	naming		
IV	symbol	color word printed in conflicting color	conceptual	reading		
Vocabulary	symbol	verbal utterance	conceptual	definition	—	
Affect-focused interview	symbol	interviewer's behavior	conceptual	affective utterance		—

Because of the critical significance of emotional changes in mental illness, it seems important that base lines of autonomic response patterning should be developed for the various psychiatric groups. Malmo (29) has recently studied pathological anxiety in this way. It may be possible to develop indices which will constitute objective correlates of global clinical concepts such as "flattened affect," "degree of tension," "process and reactive schizophrenia," etc. Such an achievement would be a considerable contribution to the prediction of outcome in schizophrenia and might aid in distinguishing clinical subgroups within the psychosis. By studying indices of autonomic reactivity which reflect the readiness of the organism to adjust to change, one may gain insight into some of the determinants of the duration of illness.

Perceptual Techniques

The significance of perceptual behavior for prognosis becomes more apparent when the overlay of conceptual activity is peeled off. Moreover, techniques which minimize the conceptual component, permit a comparison between the patient's perceptual and conceptual function. As intimated above, it may be the difference between these levels rather than their absolute function which distinguishes those who get well from the chronics.

3. *Weight judgment* (*method of absolute judgment*). This is a test of perceptual behavior in response to configural stimulation of varying complexity. The patient is given five weights to judge according to a range of categories from very heavy to very light. After the accuracy of judgment has been determined, a heavy anchor is introduced just before the presentation of each standard weight. The introduction of the heavy anchor tends to make the succeeding standard weight appear lighter. The *S* is instructed to judge only the second or standard weight. In this way, the directions attempt to offset the influence of the anchor. The effectiveness of the instructions is measured by the amount of shift in the judgment categories following the introduction of the anchor.

In a recent study, Salzinger (36) found that patients did not differ significantly from normals in accuracy of weight judgments, but did tend to shift significantly more than the normals, i.e., the schizophrenics were less successful in resisting the influence of the heavy anchor.

4. *Critical flicker frequency.* The power of the visual system to resolve stimuli closely spaced in time is a stable measure for the individual under basal conditions but is altered by damage to the central nervous system, by drugs, by illness, and by the normal process of aging (20, 24). There is some preliminary evidence that outcome of illness for schizophrenic patients may be related to level of cff at time of admission. In two studies, patients who were eventually discharged had *lower* thresholds at admission than those who remained in the hospital. As anxiety serves to reduce cff (21), it is a tenable hypothesis that patients who show greater anxiety, and hence lower cff, at time of admission have a better prognosis.

5. *Necker cube.* The number of fluctuations in the perception of an ambiguous figure, such as the Necker cube, can be counted by having the *S* press a switch every time he sees the cube shift. The data may then be analyzed for the number of shifts in each minute as well as for the total number of shifts.

Eysenck (7), following the lead of McDougall and of Hunt and Guilford, found that schizophrenics had a lower rate of reversals than normals. It is not surprising, then, as a recent study at Brooklyn State Hospital has disclosed, that first admission mental patients who twelve months later were out of the hospital had a higher rate of shift than those who remained in the hospital. This relationship, however, was reversed in a readmission group—with the "Ins" having a higher number of shifts than the "Outs." All together, the first admission Ins had the lowest number of shifts, and the readmission Ins had the highest number of shifts. Moreover, sicker patients reached a maximum rate of reversals earlier in the five minute observation period.

6. *Archimedes Spiral.* A rotating Archimedes Spiral will be seen as a contracting or expanding configuration, depending on the direction of rotation. When the rotation is stopped, a contrary after-effect is induced in normals, but not in some brain-injured patients (34). This technique is being introduced with the expectation that it may help to detect brain-injured cases among the functionally ill.

7. *Digital stimulator.* It has been known for some time that perceptual defects may be uncovered by use of simultaneous stimulation methods (13). Although little work has been done on schizophrenic patients, reports thus far available indicate that schizophrenic children show defects in tactual perception with the Face-Hand test. Adult schizophrenics also show perceptual defects when tested by simultaneous stimulation methods. A test apparatus has been designed to permit simultaneous stimulation of one to eight fingers at a constant intensity of stimulation. Preliminary analysis indicates that the amount and pattern of errors made by schizophrenic patients are different from those of matched controls.

8. *Color form test.* Thurstone's color-form test was developed for the objective measurement of color versus form preference. It consists of a moving-picture film in which the *S* indicates the direction of the apparent movement noted in the presented figure. If he sees the movement in one direction, he is color-dominant, if in the other, he is form-dominant. Some preliminary work with this technique indicates that depressed schizophrenics prefer the form responses and the downward-tending direction more often than do nondepressed patients.

Psychomotor Techniques

It has long been noted that mental patients differ from normals in their motor behavior. Indeed, the general apathy and sluggishness of many mental patients is one of their most striking characteristics, while at the other extreme the intense motor activity of the manic is often the first sign of abnormality. Recently, considerable evidence has accumulated (16) that psychomotor behavior may serve to distinguish subgroups among patients. It is in the hope of finding discriminators that will separate those who will eventually get well from the patients who fail to improve, that techniques 9 through 11 have been introduced.

9. *Purdue pegboard.* This is a standard test (44) of finger dexterity with a great deal of data on normals available for comparison. In recent studies, chronic patients performed more poorly than new admissions. Among the new admissions, significant differences were found between patients who were being hospitalized for the first time and those who had records of previous hospital admissions for mental illness.

10. *Tapping test.* A large number of studies have indicated a general slowing and an increased variability in choice reaction time in schizophrenic patients (16). Recent work has

tended to confirm these results and in addition patients with a good prognosis have been found to improve with successive testing, while those with poor prognosis showed no consistent trend. On the other hand, the trends reversed on this test for first admission patients and readmission patients. First admissions who performed well had a good prognosis, but in the readmission group poor performance was associated with good outcome. In general, the same trends hold for total tapping time as for choice reaction time, although the test seems to be somewhat less sensitive in distinguishing between groups.

11. *Reaction time for sensory modality shift.* This task requires the *S* to make a choice reaction to each of four stimuli, two in the visual modality and two in the auditory modality. These four stimuli are presented in a random series and the reaction times are classified into two groups: those for which the stimulus to the previous reaction is in the same sensory modality as the reaction under scrutiny; and those for which the stimulus to the previous reaction is in the other sensory modality. This technique is still in the pilot stage, but it is expected that schizophrenics will have a greater lengthening in reaction time than normals when there has been a shift from one modality of stimulation to the other.

Conceptual Techniques

The domain of conceptual behavior has for so long been the favorite habitat of the clinician that it hardly requires any justification, for it is indeed precisely at this level of functioning that mental disease manifests itself. Perhaps as a natural consequence, a confusion between categorization and description has vitiated much of the research at this level. In order to derive reliable evidence of conceptual functioning both for diagnosis and, more importantly, for prognosis, the psychopathologist must turn from the easy way of global observation up the stony path of rigorous measurement. Techniques numbered 12 through 19 represent some familiar and some new attempts to introduce quantification into the study of conceptual behavior. Here again the imposition of a stress-producing load helps to magnify the differences between patients and normals.

12. *Delayed auditory feedback.* The alterations in speech resulting from an experimentally induced delay in auditory feedback have been extensively studied. Individual reactions to delayed feedback vary from almost no disturbance to complete blocking of speech. The extent of the interference may represent the degree to which the individual makes use of air-conducted auditory feedback to steer ongoing speech. Experimental variation of the input to this steering mechanism provides a unique opportunity to study such processes in psychiatric populations. In a pilot study, some schizophrenics showed less disturbance than normals under delayed auditory feedback. These results corresponded with a preliminary study by Goldfarb (10) on schizophrenic children. Recently, normative and developmental studies were begun to obtain baselines for studies of abnormal populations. Preliminary results indicate substantial differences between normal children aged 4 to 6 and normal children in the 7- to 9-year age range.

13. *Metalog.* The Metalog test (46) attempts to tap facility of conceptual shift by requiring the *S* to produce as many different meanings to a simple stimulus word as he can within one minute. A high score corresponds to a large number of different meanings. This test was first used in connection with the Columbia-Greystone (25) study where a high score was found to be associated with poor prognosis in both the operated and unoperated groups.

14. *Analogies test.* In this test of conceptual performance, the *S* is required to complete a series of verbal analogies of the type, "Night is to day as dark is to *light.*" Like the Metalog test, this test too was found to be inversely related to outcome for patients in the Columbia-Greystone study (25).

15. *Essential differences test.* In this test, the *S* is presented with successive sets of four words and is required to indicate in each set which word is most different from the other three; for example, apple, pear, *carrot*, plum. The *S* thus has to abstract an attribute which is common to three of the four words (25). He is also asked to justify his selection, i.e., to give an account of the ideational process involved. This is one of several tests which seem to be positively related to outcome for first admissions, but negatively related to outcome for readmissions.

16. *Porteus maze.* A standard test of nonverbal conceptual functioning, the Porteus (33) invokes foresight and planning. This test has shown significant changes after brain surgery.

Chronic schizophrenic patients tend to score quite poorly on this test.

17. *Stroop test.* This test (43) compares the ability to perform a conceptual task involving discrimination of four colors with and without distraction. The *S* is first asked to read a list of color names. His next task is to assign the proper names to a random sequence of colored dots. Then he is presented with a list of color names printed in a conflicting color of ink. He must name as rapidly as possible the color of the ink in which the words are printed. Finally, he is told to ignore the actual colors and to read the color names as fast as he can. The difference between the speed with which colors are named in the conflict situation and the speed when no conflict is present provides a critical score. Ability to resist the distraction of the conflict card appears to be positively related to outcome of schizophrenic illness.

18. *Vocabulary test.* Vocabulary can be considered a good indicator of overlearned conceptual behavior for purposes of comparison with other measures of conceptual functioning. The word list is taken from the Wechsler and Binet tests.

19. *Affect-focused interview.* Distortion of affective experience is one of the major characteristics of mental disorder. In fact, patients seldom appeal for help because of any inadequacy in adjustment to the environment. They usually come to the physician because of the way they feel. Moreover, one of the major findings of a survey of the literature was that patients with flattened affect had poor prognoses. In the affect-focused interview (12, 37, 45), the patient is first engaged in conversation for ten minutes (the operant period). During the second ten-minute interval, every affective utterance is reinforced by expression of assent, such as "umhum" (conditioning period). In a third ten-minute interval, the reinforcement is withdrawn (extinction period). An analysis of recorded interviews for the number of affective utterances during the operant, conditioning, and extinction periods gives a measure of the ability of the patient to emit affective utterances under reinforcement. The procedure has been refined by controls for reliability, i.e., repeated interviews with different interviewers, and repeated interviews on different occasions with the same interviewer. The interview is conducted in a room provided with a one-way observation screen so that the nonverbal behavior of interviewer and interviewee can be observed.

SUMMARY

The aid of biometrics is invoked for the design of a program of experimentation; for the integration of findings at different levels of complexity; and for the development of new research tools.

The behavior of individuals can be analyzed into a set of overlapping domains of ascending psychological complexity: physiological, sensory, perceptual, psychomotor, and conceptual. These areas have not received equal attention from students of psychopathology. In the past, there has been too much preoccupation with the conceptual domain. In fact, a good deal of research labeled as perceptual is actually concerned with conceptual behavior. While sensory psychology has thus far contributed little to our understanding of psychopathology, the neglect of physiology by most psychopathologists is unfortunate. Moreover, the interrelations of these functional levels may prove to be of prognostic value. Thus, there is already some evidence that the ratio of perceptual to conceptual function distinguishes different subcategories of mental patients.

The variety of investigative methods employed in research ranges from naturalistic observation through controlled experiments to modifications of the organism. The first method, observation, is propaedeutic to science. The laboratory experiment involves directing energy or signals at the organism and studying that part of the subsequent behavior which can be functionally related to the independent variable. Variables of the energy category may be appropriate (e.g., light for the eye) or inappropriate (e.g., pressure on the eyeball). Signal variables can be: simple configurations (as in studies of perceptual discrimination); signs which represent some event remote in time or space (as in learning experiments); or symbols (as in studies of conceptual function.)

In order to throw additional light on the behavior of the intact organism, it is often useful to study response to controlled stimulation when the subject is under the influence of a stress-producing load. By straining the organism to its limit of resilience, a load may help to reveal differences between normals and patients which are obscured under everyday conditions.

Radical invasions of the organism's integrity, such as administration of hallucinogenic drugs, implantation of electrodes or ablation of brain tissue, can provide information about limited subsystems of behavior, but raise many problems of relevance with respect to integrated behavior.

A comprehensive program of research is presented, part of it already under way, designed to exploit the possibilities inherent in the interactions between different levels of psychological function.

REFERENCES

1. BARD, P., & MOUNTCASTLE, V. B. Some forebrain mechanisms involved in expression of rage with special reference to suppression of angry behaviour. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, **27,** 362–404.
2. BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Can. J. Psychol.*, 1954, **8,** 70–76.
3. BREMER, F. Dualité des processus d'excitation centrale. *Ann. Physiol. Physiochim. biol.*, 1933, **9,** 897–903.
4. BURDOCK, E. I., & ZUBIN, J. A rationale for the classification of experimental techniques in abnormal psychology. *J. gen. Psychol.*, 1956, **55,** 35–49.
5. CANNON, W. B. *Bodily changes in pain, hunger, fear, and rage.* New York: Appleton-Century-Crofts, 1939.
6. DELGADO, J. M. R., HAMLIN, H., & CHAPMAN, W. P. Technique of intracranial electrode implacement for recording and stimulation and its possible therapeutic value in psychotic patients. *Confin. neurol.* (*Basel*), 1952, **12,** 315–319.
7. EYSENCK, H. J. Schizothymia-cyclothymia as a dimension of personality. *J. Pers.*, 1952, **20,** 345.
8. FRITSCH, G., & HITZIG, E. Ueber die elektrische Erregbarkeit des Grosshirns. *Arch. Anat. Physiol. wiss. Med.*, 1870, **37,** 300–332.
9. GALAMBOS, R. Neural mechanisms of audition. *Physiol. Rev.*, 1954, **34,** 497–528.
10. GOLDFARB, W. Reactions to delayed auditory feedback among a group of children institutionalized for severe behavior disorders. In P. H. Hoch & J. Zubin (Eds.), *Psychopathology of communication.* New York: Grune & Stratton, in press.
11. GRANIT, R. Centrifugal and antidromic effects on ganglion cells of retina. *J. Neurophysiol.*, 1955, **18,** 388–411.
12. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68,** 409–416.
13. HEAD, H., & HOLMES, G. Sensory disturbances from cerebral lesions. *Brain*, 1911, **34,** 102–254.
14. HOFFER, A., OSMOND, H., & SMYTHIES, J. Schizophrenia: A new approach. II. Result of a year's research. *J. ment. Sci.*, 1954, **100,** 29–45.
15. KEYS, A., BROZEK, J., HENSCHEL, A., MICKELSEN, O., & TAYLOR, H. L. *The biology of human starvation.* 2 vols. Minneapolis: Univer. of Minnesota Press, 1950.
16. KING, H. E. *Psychomotor aspects of mental disease.* Cambridge: Harvard Univer. Press, 1954.
17. KLÜVER, H. An experimental study of the eidetic type. *Genet. Psychol. Monog.*, 1926, **1** (No. 2), 69–230.
18. KLÜVER, H. *Mescal: The divine plant and its psychological effects.* London: K. Paul, Trench & Trubner, 1938.
19. KLÜVER, H., & BUCY, P. C. Preliminary analysis of functions of temporal lobes in monkeys. *Arch. Neurol. Psychiat.*, 1939, **42,** 979–1000.
20. KRASNO, L. R., & IVY, A. C. The response of the flicker fusion threshold to nitroglycerin and its potential value in the diagnosis, prognosis, and therapy of subclinical and clinical cardiovascular disease. *Circulation*, 1950, **1** (No. 6), 1267–1276.
21. KRUGMAN, H. E. Flicker fusion frequency as a function of anxiety reaction: An exploratory study. *Psychosom. Med.*, 1947, **9,** 269–272.
22. KUFFLER, S. W., & HUNT, C. C. The mammalian small nerve fibers: a system for efferent nervous regulation of muscle spindle discharge. *Assn. Res. Nerv. Ment. Diseases*, Research publ. **30,** 1952, 24–47.
23. LACEY, J. I., BATEMAN, D. C., & VAN LEHN, R. Autonomic response specificity: An experimental study. *Psychosom. Med.*, 1953, **15,** 8–21.
24. LANDIS, C. Determinants of the critical flicker fusion threshold. *Physiol. Rev.*, 1954, **34** (No. 2), 259–286.
25. LANDIS, C., GARRISON, M., HAMWI, VIOLET, STAUFFER, ANNE K., KING, H. E., & YOUNG, KATHLEEN M. Test results one year after operation. In F. A. Mettler (Ed.), *Selective partial ablation of the frontal cortex.* New York: Paul B. Hoeber, 1949. Pp. 302–305.
26. LASHLEY, K. S. Dynamic processes in perception. In J. F. Delafresnaye (Ed.), *Brain mechanisms and consciousness.* Oxford: Blackwell, 1954. Pp. 422–443.
27. LEHRMAN, D. S. A critique of Konrad Lorenz's theory of instinctive behavior. *Quart. Rev. Biol.*, 1953, **28,** 337–363.
28. LOWENSTEIN, O. Pupillography. *AMA Arch. Ophthal.*, 1956, **55,** 565.
29. MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11,** 9–24.
30. MEAD, G. H. *Mind, self and society.* Chicago: Univer. of Chicago Press, 1934.
31. MORISON, R. S., DEMPSEY, E. W., & MORISON, B. R. Cortical responses from electrical stimulation of the brain stem. *Amer. J. Physiol.*, 1941, **131,** 732–743.
32. OLDS, J., & MILNER, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. comp. physiol. Psychol.*, 1954, **47,** 419–427.
33. PORTEUS, S. D., & PETERS, H. Maze test validation and psychosurgery. *Genet. Psychol. Monogr.*, 1947, **36,** 3–86.

34. PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.
35. RICHTER, C. P. Total self regulatory functions in animals and human beings. *Harvey Lectures*, 1942–43, **38**, 63–103.
36. SALZINGER, K. Shift in judgment of weights as a function of anchoring stimuli and instructions in early schizophrenics and normals. Unpublished doctoral dissertation, Columbia Univer., 1954.
37. SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. Paper read at East. Psychol. Ass., New York, April, 1957.
38. SCHREINER, L., & KLING, A. Behavioral changes following rhinencephalic injury in the cat. *J. Neurophysiol.*, 1953, **16**, 643–659.
39. SELYE, H. The general adaptation syndrome and the diseases of adaptation. *J. clin. Endocrinol.*, 1946, **6**, 117–230.
40. SIDGWICK, H. *et al.* Physiology of hallucinations. *Proc. Soc. psychical Res.* (*E.*), 1894, **10**, 134–148.
41. SMYTHIES, J. R. *Analysis of perception.* London: Routledge & Kegan Paul, 1956.
42. STOLL, W. A. Lysergsäure-diäthylamid ein Phantasticum aus der Mutterkorngruppe. *Schweiz. Arch. Neurol. Psychiat.*, 1947, **60**, 1.
43. THURSTONE, L. L., & MELLINGER, J. J. *The Stroop test.* University of North Carolina, Psychometric Laboratory, No. 3, 1953.
44. TIFFIN, G., & ASHER, E. J. The Purdue pegboard: Norms and studies of reliability and validity. *J. appl. Psychol.*, 1948, **32**, 234.
45. VERPLANCK, W. S. The control of the content of conversation by reinforcement. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.
46. ZUBIN, J., & WINDLE, C. The prognostic value of the Metenym test in a follow-up study of psychosurgery patients and their controls. *J. clin. Psychol.*, 1951, **7**, 221–223.
47. ZUBIN, J., WINDLE, C., & HAMWI, VIOLET. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, **21**, 342–355.

Received January 30, 1957.

THE PSYCHODYNAMICS OF SOCIAL MOBILITY IN ADOLESCENT BOYS

ELIZABETH DOUVAN AND JOSEPH ADELSON
Survey Research Center, University of Michigan, and University of Michigan

There is a large and growing body of literature on social mobility, and the significance of the topic is generally acknowledged; so it is surprising to find that only limited attention has been given to studying the motivational sources of mobility. What we do find is a general disposition to treat *upward* mobility in a vaguely invidious fashion. It would seem that, in this country, the Horatio Alger tradition and the "dream of success" motif (6) have been pervasive and distasteful enough to have alienated, among others, a good many social scientists. The upwardly aspiring individual has apparently become associated with the pathetic seeker after success or with the ruthless tycoon. This image of success is, much of it, implicit—assumption and attitude, and not quite conviction—but it seems to have dominated the thinking of our intellectual community.

Or so it has been until recently. Newer empirical findings have encouraged a more differentiated view of mobility. We begin to get some sense that varying motivations may underlie social striving; we become aware that the direction, the rapidity, and the absolute extent of mobility may reflect differing psychodynamic sources and thus require separate analytic treatment.

In this paper, a first effort is made to explore and clarify one problem in this complex area—to analyze some of the psychic accompaniments of upward and downward mobility strivings among adolescent boys. In the following section, a theory of the personality determinants of mobility is presented; the paper as a whole is devoted to testing some of these formulations.

THEORETICAL CONSIDERATIONS

Upward Mobility

We can distinguish at least three separate patterns of motivation which are implicit in discussions of upward social mobility: fear of failure, ambivalence toward success, and hope of success.

The first pattern—perhaps the most common image of the mobile individual—is of the desperately scrambling Philistine, exhausted in his pursuit of status. He has defined his identity and personal worth exclusively by the criteria of success. Perhaps a typical genetic source of this pattern has been described by Ackerman and Jahoda (1). Competitiveness is taught by the family: "Most of the mothers of our cases . . . apparently did not tell their children 'be happy' but rather 'make money'. . . ." "Success is measured by comparison with others rather than by actual achievement. . . . There are always some who have done better, who have more money and more social prestige; and there is always the danger of being pushed down the social ladder by a competitor" (1, pp. 88–89).

In the second pattern—ambivalence toward success—the dynamics are more complex. The individual aspiring upward is here seen as responding to the dominance and prestige of the more privileged stratum through a defensive identification with this group and by denying and decrying his own background and status. The genetic sources of this choice of mechanism may be sought in the aspirer's early encounters with authority and status in the original family setting. We discover in the family scene a harsh and forbidding parent, one who allows the child no hostility; we find a child who has accommodated to parental strength through identification. In this formulation, we see the mobile individual showing many of the characteristics of the "authoritarian personality": a conscious over-idealization of the parent and of authority together with unconscious rebellion and hostility, rigidity, conformity, and anti-intraception. As regards mobility, we find the following dynamics: the motive for upward aspiration arises out of the need to emulate the parents; at the same time mobility arouses conflict since it implies a struggle with this powerful authority figure (2).

Without question, both of these mobility types occur (perhaps frequently) in our society. Yet the recent literature on mobility has suggested that the most common pattern is the one we have called "hope of success" (7). Here we find interest in success directed by a rational ego: the individual who can mobilize his energies effectively, whose aspirations are realistic. The mobility goals are moderate or, in those cases where aspiration is toward a substantial status move, they are nevertheless realistic in view of the individual's talents.

We find a family milieu which encourages autonomy, is not obsessed by status, yet accepts and transmits without ambivalence the culturally central value of achievement. Moreover, the shifts in identity required do not implicate the child in conflict: the family has provided or permitted the child a model for identity consistent with the goal of mobility.

In a society which holds upward mobility to be a central value, which provides opportunity for it, and in which it is, apparently, a very common occurrence, we can expect that the dominant motivational pattern informing upward aspiration is not a defensive one—does not necessitate a personally damaging flight from one's past; rather, we would expect it to accompany effective ego functioning and successful socialization by a family which is transmitting a cultural value of which it approves.[1]

Downward Mobility

What about the adolescent boy whose aspirations are downward in direction? Here we have someone whose goals are atypical; indeed, they contradict directly a strongly held cultural value. To be sure, the same novelistic tradition which associates upward mobility with Babbittry has sometimes offered us a sentimental image of the downward mobile type, presenting him as an unfettered, romantically sullen child of nature. Our own view is considerably darker.

We see downward aspiration as representing, in many areas, the psychodynamic opposite of upward mobility. If we are correct, these adolescents are demoralized, alienated, anomic. An ambivalent relationship to the parents produces an impoverished ego, vulnerability to conflict, a failure to internalize general cultural values. Ambivalence is, of course, an almost universal outcome of the socialization process; but we conjecture that it is especially acute among the downward mobile. We expect to find a relationship to the parents which is infantile, that is, dependent and hostile. Dependency and aggression accompany, reinforce, and conceal each other. The child who is captured in an infantile tie to his parents has too much of his energy committed in ambivalence; ego functions remain immature and conflict-ridden; the ego is unable to cope adequately with impulses, internalized morality, and the more complex aspects of reality. At its best, the ego enforces a harmony among the psychic institutions and the outside world. But the downward mobile boy, if our formulation is correct, is at the mercy of his impulses, which continually threaten to overwhelm brittle controls. The superego drives are equally primitive and frightening; parental prohibitions are not sufficiently internalized, in part because of their very intensity; superego functioning is consequently immature, in that there is an inadequate fusion of ego and superego processes; the superego stands at some distance from the ego, so that questions of morality are not so much matters of right and wrong as of escape and pursuit. As regards the ego's operations vis-a-vis the outer world, we do not hypothesize that the downward mobile show any gross pathology of reality testing. And yet we feel that the more complex and articulated ego processes—judgment, rapport, control, time-binding—are poorly or unevenly developed.

Hypotheses

In the research reported here, a study of the mobility aspirations of adolescent boys, we test a number of hypotheses based on these conceptions of the sources of mobility strivings. Specifically, the following predictions were made about upward- and downward-aspiring boys.[2]

---

[1] One may speculate about those social situations in our culture where upward aspiration necessitates a good deal of emotional stress or induces personal conflict. Generally, we would say, these are situations where a change in status requires a decisive modification of behavior or necessitates a sharp shift in identity models. A number of such situations come to mind: (*a*) where the status change involves a concomitant ethnic (including religious) shift; (*b*) where it involves a gross movement up the status ladder; (*c*) where there has been a limited opportunity to learn the status behavior of the higher status group; (*d*) at certain points in the status continuum, where the defining criteria of the higher status group are realistically difficult to achieve; e.g., in the movement from lower-upper to upper-upper status, in those cases where a defining criterion is ancestry; (*e*) where the group-of-origin is ambivalent or hostile to the mobility of its members, necessitating a decisive abandonment of the past as a condition of mobility. Under these conditions, mobility upward will require an unusually high degree of personal motivation.

1. *Energy level.* The upward-mobile boys possess a high degree of available energy for use in social and work activities. Downward-mobile boys, on the other hand, show a diminished vitality.

2. *Achievement mode.* The boys aspiring to higher status show a pervasive achievement orientation and a secure sense of their own effectiveness in reaching goals. The downward mobile are less oriented toward achievement and dominated by a concern with security.

3. *Orientation toward the future.* The time perspective of the upward mobile tends to be extended, while the downward mobile tend to have a constricted orientation in time.

4. *Personal standards.* The upward mobile boys manifest well internalized moral values and standards of personal behavior. Internalization is relatively incomplete among the downward mobile.

5. *Autonomy.* The upward aspiring display a precocious independence from the family, and a strong sense of autonomy in choosing values and goals. Downward mobile boys, on the other hand, are tied in a dependent relationship to the family.

6. *Orientation toward the self.* Upward-mobile boys are realistic in assessing themselves, and show a high degree of self-confidence. The downward aspiring reveal ambivalence toward the self and a lack of poise in social situations.

7. *Family milieu.* In this area, our predictions were firmer and more specific about the downward mobile group. Previous research (3) led us to expect a pattern of ambivalent dependency in the downward mobile boys' relations with their families. Rebellion and rejection of parental values combined with a strong dependency stem from inconsistent and punitive treatment at the parent's hand: in short, we anticipated an authoritarian family setting.

In addition to these specific predictions, we were interested in inquiring whether the relationship to the family, in upward mobile boys, is characterized by an ambivalent rejection of family values or by a conflict-free differentiation of the self from the family.

METHOD

Subjects

The data for our analysis derive from a national sample survey of adolescent boys conducted by the Survey Research Center.[3] The sample for the total study consisted of a thousand boys in the 14- to 16-year age range selected by probability sampling methods. Each boy was given a personal interview at school by a member of the research center's field staff. Open-ended questions and projective questions were used, and interviews lasted from one to three hours.

In the present analysis, we have used as the base sample the interviews taken with urban nonfarm youth whose fathers' occupations fell in any of the following categories: small business owners, self-employed artisans, white collar, sales, and clerical workers; skilled and semi-skilled manual workers. The reason for excluding the sons of men occupying positions at the two extremes of the skill hierarchy becomes apparent when the mobility measure is discussed.

Mobility Aspiration

Each of our youthful respondents was asked the following questions:

a. What kind of work would you like to do as an adult?

b. Are you pretty sure about this or do you think you're just as likely to go into something else?

Each boy's aspiration was classed on the occupational scale and compared to his father's position to determine whether it was equivalent to or higher or lower than the father's job in the hierarchy of skills and status.

The distribution of aspiration types by father's occupation is presented in Table 1 for

---

[2] In this paper we will use the terms "upward aspiring" and "upward mobile" interchangeably. It should be clear that we do not assume that upward aspiration will necessarily lead to successful social mobility.

[3] The survey was sponsored by the Boy Scouts of America. The complete questionnaire used in the study has been deposited with the American Documentation Institute. Order Document No. 5422 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for 35 mm. microfilm or $2.50 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

DISTRIBUTION OF ASPIRATION TYPES BY FATHER'S OCCUPATION

	Father's Occupation			
Aspiration	1 Professional, Managerial ($N = 136$) %	2 White Collar[a] ($N = 183$) %	3 Manual Skilled[b] ($N = 335$) %	4 Manual Unskilled ($N = 70$) %
Upward		46	58	84
Stable	67	29	34	16
Downward	33	25	08	

Note.—Ten per cent of the total urban sample of boys was lost in the process of rating the boy's aspiration in relation to father's occupation. These were *S*s who had not decided what they wanted to be, who gave vague answers, or who wished to be farmers.

[a] This category includes small business owners in addition to sales and clerical workers.

[b] This category includes both skilled and semi-skilled workers.

all urban youth in the original sample. An obvious fact is illustrated in this table. Boys whose fathers currently occupy jobs in the top category are by definition limited to equivalent or lower status aspirations, just as those whose fathers are unskilled workers are barred from downward aspiration.

Because of this limitation on the relative freedom of direction in the aspirations of boys from families in Groups 1 and 4, and because we were interested ultimately in both upward and downward mobility, the sample was restricted to those boys from the two middle categories. The aspiration behavior of boys from the extreme groups may reflect in part a statistical artifact. That is, assuming that by chance a certain proportion of boys will aspire to a status-skill level different from the parent's, the direction of deviation is automatically determined for boys in Groups 1 and 4. In the middle groups, on the other hand, deviation is possible in either direction, and the selection of one or the other type should reflect certain specific motivational factors. One might maintain that the fact of aspiring to a level other than the father's in itself reflects unique psychological characteristics, but since the direction of deviation is a focus of our present concern, opportunity for choice in both directions was controlled.[4]

RESULTS

In the following pages we show contrasts among upward, downward, and stable groups. In most cases, we find linear relationships between status aspiration and other variables: the upward mobile are particularly high or low on a variable, the downward mobile are at the opposite extreme, and the stable group falls somewhere in between. Occasionally we find that either the upward or downward aspiring group clearly distinguishes itself from the other two groups. And in a few relationships, the stables stand out, the linear relationship gives way to a curvilinear one.

Since our predictions all concern the mobility categories and there was uncertainty about the exact position of the stable group, chi square was used to test the significance of variation among the three groups.

Energy Level

We see here a sharp contrast between upward and downward subjects (*S*s): while the upward mobile boy is unusually lively and energetic, the downward aspirer is inactive and apathetic. On all of our measures, the upward group reveals a high commitment of energy to social and recreational pursuits, and yet not at the expense of work involvement; they also seem to show a spirited and enthusiastic love of activity. The downward mobile boy, in contrast, shows an impoverished vitality: he does not participate energetically or enthusiastically in recreational activities.

Looking first at boys' memberships in organizations (Table 2, Item 1), we find that three quarters of the upward mobile *S*s belong to some organized group; while slightly less than half of the downward aspiring boys report membership. The upward mobile also have a relatively high proportion of multiple memberships, compared to the other two groups. A similar relationship appears in the area of general leisure activities: almost half the upward mobile report having tried twenty of more separate activities while only about one quarter of the downward mobile fall in

---

[4] The group used for analysis consisted of 335 boys from working-class background and 183 from white-collar families. Since the upward mobile group contained a disproportionate number of boys from blue-collar backgrounds, we ran all mobility analyses separately for the two broad occupational groups (Categories 2 and 3 in Table 1). Relationships were of approximately the same order within each of these groups. Analysis is presented for the combined group in order to avoid complicating tables unnecessarily, but in all cases relationships reported also held within each of the two background groups.

TABLE 2

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF LEVEL OF ENERGY

Item	Mobility: Upward (N = 277) %	Mobility: Stable (N = 168) %	Mobility: Downward (N = 73) %	χ^{2*}	p Level
1. Number of group memberships[a]					
a. none	25	33	52	26.91	.01
b. one	32	38	22		
c. two	23	20	17		
d. three or more	20	09	09		
2. Number of leisure activities					
a. fewer than 20	52	65	67	11.65	.01
b. 20 or more	48	35	33		
3. Dating					
a. do date	66	59	52	5.67	.10–.05
b. do not date	34	41	48		
4. Employment					
a. have jobs	49	47	51	.44	.90
b. do not have jobs	51	53	49		
5. Leisure reading					
a. do not read	17	27	25	7.93	.05
b. fiction-novels, mysteries	32	23	11	12.31	.01
c. travel and adventure	31	20	18	10.52	0.1
d. technical, scientific	09	04	01	11.13	.01
e. sports and hobby books	10	09	04	2.44	.30–.20
f. history, biography	08	07	05	.27	.90–.80
g. animal stories	05	04	05	.73	.70–.50
h. comics, joke books	21	26	27	1.95	.50–.30
i. newspapers, magazines	19	16	19	.96	.70–.50
6. Proportion of activities enjoyed					
a. fewer than half	45	41	59	6.09	.05
b. one half or more	55	59	41		
7. Additional activities desired					
a. suggest activity	63	31	10	87.67	.01
b. do not suggest activity	37	69	90		

[*] Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] In this case, the chi square was derived from a 3 x 4 table with six degrees of freedom.

this category; the stable group is approximately midway between the other two (Table 2, Item 2). There is a small difference in dating, the upwards being slightly more likely, and the downwards least likely to date (Table 2, Item 3).

Although they more commonly engage in leisure pursuits, the aspiring boys are not significantly different from others in having job commitments. One might expect that the reason the downward-aspiring boys play less is that they work more. But this does not seem to be the case (Table 2, Item 4), the three groups being approximately equal in the percentage who hold jobs. The more active social life of the upward mobile boy represents an increment in total activity rather than a substitute for work activity. The upward mobile boy is more likely to report leisure reading. He has a more diversified reading pattern, and his reading includes more demanding material. The downward aspiring boy much less often reads novels and mysteries, travel and adventure stories, and technical, scientific literature. Stable boys fall between the other two groups on all specific categories of reading, but are like the downward mobile in the proportion of nonreaders. The single case in which the trend favors the downward mobile boy is in the category "comics, joke books" (Table 2, Item 5).

Upward mobile boys are not only more active currently but also seem to be more enthusiastic about their activities and more receptive to new experiences. After our *S*s had checked the leisure pursuits they had experienced within the last year, they were asked to check the ones they had especially enjoyed. In general, the two indices bear a low negative relation to each other—that is, boys who do more things are more selective in naming those they "particularly enjoy." Despite this tendency for a negative association, we find that the upward aspiring boys (who take part in more leisure activities) are nevertheless enthusiastic about the things they do. They are like the stables in the proportion of their leisure pursuits that they designate as particularly enjoyable; while the downward mobile, who have tried relatively few leisure activities, are much less enthusiastic about these few (Table 2, Item 6). When asked if there are other activities they would like to try, aside from those appearing on the list, almost two-thirds of the upward group name something else they would enjoy. This responsiveness contrasts sharply to the other groups: a third of the stable group, and only one tenth of the downward group suggest any additional unlisted activities that appear to them (Table 2, Item 7).

Achievement Mode

We expected that the degree of striving reflected in a boy's job aspiration would be part of a general achievement orientation, that upward occupational aspiration would be associated with the pattern of behavior which McClelland *et al.* (7) have termed the "achievement syndrome," and that the upward mobile boy would show concern with achievement as against security and would be interested in meeting self-imposed standards of excellence. The downward aspiring boy, on the other hand, should show the same lack of vigor, the same constriction in setting other goals that we have seen in his choice of a future occupation.

The most direct evidence on these points is found in response to other questions in the occupation area. The *S*s were asked to choose among criteria for judging the attractiveness of a job, to choose the two most and the two least important dimensions. Compared to the stable and downward groups, upward mobile boys stress interesting work and have a less dominating concern about security. They are highest of all three groups in desire for status achievement. The downward mobile also strongly wish for status, but their status drive is not bound to a demand for interest in the content of the job; rather, they would most like to have status combined with security. The answers of our stable group are interesting on this item; relatively unattracted by status and fame, they stress security and interesting work (Table 3, Item 1).

TABLE 3
RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF ACHIEVEMENT MODE

Item	Mobility: Upward (*N* = 277) %	Mobility: Stable (*N* = 168) %	Mobility: Downward (*N* = 73) %	χ^{2*}	*p* Level
1. Job criteria selected[a]					
a. interesting work	56	51	33	15.74	.01
b. status achievement	44	28	36	22.15	.01
c. security	50	59	65	5.84	.10
2. Reasons for job choice[a]					
a. status	20	09	03	11.65	.01
b. ease of job or ease of obtaining job	04	12	32	15.27	.01
c. interest of work	70	78	58	5.99	.05
3. Preference: success or security					
a. success	52	35	33	12.00	.01
b. security	48	65	67		
4. Savings					
a. do save	30	21	18	6.16	.05
b. do not save	70	79	82		
5. Education plans					
a. beyond high school	69	42	12	140.06	.01
b. not beyond high school	31	58	88		

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] Categories here are not exhaustive. Tests are run on the presence or absence of particular answers.

When we ask boys to rationalize their own job choices, the upward aspiring *S* emphasizes interest in the work itself, while the downward mobile more often adduces the ease of the work, or the ease of obtaining employment in the field (Table 3, Item 2).

It would seem, then, that the upward mobile *S*s are more willing to yield security for the sake of achievement. In answer to a forced-choice question: "Which would you rather have—a job where you're sure you won't be laid off or one you can't be so sure of but where you have a chance to be a big success?" the striving boys choose the challenging, less secure job significantly more often than boys in either of the other groups (Table 3, Item 3).

Achievement orientation is often inferred from the willingness to forego immediate gratification for long-term goals. One indication of this appears in answer to the question, "What do you spend your own money for?" The boys who aspire high are more likely to say they save their resources for education or other long term goals (Table 3, Item 4). As we would expect, boys aspiring to higher occupations also reveal more extensive educational plans than do other boys (Table 3, Item 5).

Orientation toward the Future

The upward mobile boys have an extended time perspective and a marked interest in attaining adult status; while the downward group shows constriction in time, as in other aspects of the life space. We asked our *S*s what decisions they would have to make in the next few years; the answers were rated for the breadth of time perspective. The upward

TABLE 4

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF ORIENTATION TOWARD THE FUTURE

Item	Mobility			χ^{2*}	p Level
	Upward	Stable	Downward		
	(N = 277) %	(N = 68) %	(N = 73) %		
1. Time perspective on decisions[b]					
a. within high school	17	18	29	5.21	.10
b. beyond high school	78	73	77	1.75	.50
c. distant future	53	51	26	16.51	.01
2. Sources of self-esteem[b]					
a. assuming adult role	31	23	16	6.91	.05
b. belonging, being part of a group	46	56	32	11.34	.01
c. being accepted by peers	03	07	14	10.90	.01
d. nothing	04	08	11	6.39	.05
3. Preference for coed or all-boy club[a]					
a. all boy	39	48	59	10.60	.05
b. neutral	23	16	14		
c. coed	38	36	27		

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] In this case, the chi square was derived from a 3 x 3 table with four degrees of freedom.

[b] Categories here are not exhaustive of responses given to open questions, and more than one response was often given. Tests are run on the presence or absence of particular answers.

aspiring boys are more likely to mention decisions in the more distant future, such as adult roles and goals. The downward mobile more often name immediate decisions, those within the high school period (Table 4, Item 1).

In answer to the question "What things that you do—at home, in school, or with your friends—make you feel important and useful?" the upward striving boy more often states instances in which he has assumed adult roles. He is less likely than the downward mobile *S* to mention peer acceptance as a source of self-esteem. An interesting finding here is that the stable group is highest of all three in choosing "belonging, being part of a group," one of the few occasions where this group does not fall somewhere between the other two. We suggest, though very tentatively, that this may reflect a concern with maintaining stable interpersonal ties.

We may also point to the apparently greater social maturity of the upward mobiles. As we have seen, they are somewhat more likely to date. They are also more eager for heterosexual social activities; when asked to choose between an all-boy or a coed social club, the upward aspiring less often prefer the former (Table 4, Item 3).

The Internalization of Personal Standards

Another point of difference between the groups concerns the interiorization of standards. Upward mobile boys consistently show a more complete internalization of personal controls. Downward mobile boys show a tendency to externalize standards and to rebel against them.

A set of projective pictures showed an adolescent boy in conflict between a promise made to his parents (to be home at a certain time) and peer pressure (to stay out later). Our *S*s were asked what the boy would do and how he would feel about it.

In all of our groups, approximately two thirds of the respondents say that the boy will go home. However, the motives attributed to the boy differ sharply. The upward mobile boys more often have the boy go home because "he promised" or because "his parents trust him," whereas the downwards say he will go because of a fear that the parents will find out and punish the transgression (Table 5, Item 1).

Regardless of their responses to these questions, all *S*s were asked: "If the boy decided to stay with his friends awhile, do you think he'd tell his parents about it later?" The upward aspiring group answer "yes" more frequently than other boys (Table 5, Item 2).

Another series of questions concerned boys' conceptions of and general reactions to rules. Here, again, the upward mobile group reveals a greater degree of internalization. When asked why a boy might break a rule, they focus on uncontrollable emergency situations and those in which a boy is old enough (i.e., responsible enough) to guide his own behavior. On the other hand, the downwards more frequently give responses fitting the categories of "rebellion against the parent," "irresistible impulse," and "when parental authority is not present" (Table 5, Item 3). When asked what kind of a rule they would never break, upward mobile *S*s more often mention those which involve responsibility to others; they less frequently say that there is no rule, or no rule that comes

TABLE 5

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF INTERNALIZATION OF PERSONAL STANDARDS

Item	Mobility			χ^2*	p Level
	Upward	Stable	Downward		
	(N = 277) %	(N = 168) %	(N = 73) %		
1. Reaction to parent-peer conflict[a]					
a. adheres to promise because of sense of trust	37	19	16	24.03	.01
b. Adheres to promise because of fear of punishment	04	06	14	6.38	.05
2. Honesty with parents					
a. would tell parents	55	42	44	9.15	.05
3. Conditions for breaking a rule[a]					
a. emergency	18	10	10	6.56	.05
b. boy mature enough	11	07	01	6.49	.05
c. rebellion	10	14	22	10.11	.01
d. impulse	05	11	16	11.41	.01
e. parental authority absent	05	10	27	31.50	.01
4. Unbreakable rule					
a. no unbreakable rule	03	17	21	10.85	.01

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] Categories here are not exhaustive of responses given to open questions, and more than one response was often given. Tests are run on the presence or absence of particular answers.

to mind, which they wouldn't break (Table 5, Item 4).

Autonomy

We saw earlier that the striving boys are interested in assuming adult-like roles, those, that is, in which they themselves are responsible for their behavior and for a job. We see here a seemingly greater drive towards independence and responsibility. In contrast, the downward mobile, though they show signs of rebellion against the parents, are also more dependent on them.

In one series of items, our *S*s were asked whose advice they would prefer to take on particular problems—their parents' or friends'. The issues ran from such central ones as what time to be in at night to matters of taste, such as personal grooming. While there were no very striking differences in the proportions who would heed their parents on any of these topics, one interesting difference appears in the proportion of boys who inject a note of independence while discussing these issues. Thirty per cent of the striving group spontaneously assert that they would follow their own ideas on at least one of the six issues, and only three per cent of the downwards give such a response (Table 6, Item 2).

We find another indication of independence from the family in response to the question, "Can a friend ever be as close as a family member?" The upward mobile *S*s most often agree to this idea (Table 6, Item 3).

When our *S*s are asked what the role of a club leader ought to be, we again find that the aspiring boys desire independent direction of their behavior. They limit the leader's function to that of a helper (Table 6, Item 4). A coder rating based on this question and on one which asked the *S* to describe an adult leader he had liked, revealed that the downward mobile *S* is

TABLE 6

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF AUTONOMY

Item	Mobility			χ^2*	p Level
	Upward	Stable	Downward		
	(N = 277) %	(N = 168) %	(N = 73) %		
1. Independence in allocating funds					
a. no independence	01	01	19	59.27	.01
2. Advice on decisions					
a. interjects own opinions	30	21	03	24.69	.01
3. Intimacy of friendship					
a. can be as close as family relationship	59	54	34	9.92	.01
4. Role of adult leader[a]					
a. helper	58	51	38	9.71	.01
b. decision maker	40	43	51	3.67	.20
5. Authority reliance					
a. high	38	47	54	6.06	.05
b. moderate	62	53	47		
6. Adult ideal[a]					
a. family member	41	50	55	5.99	.05
b. unrelated adult acquaintance and composite	31	24	18	6.74	.05

* Unless noted, the chi square was derived from a 3 x 2 table with 2 degrees of freedom.

[a] Categories here are not exhaustive of responses given to open questions, and more than one response was often given. Tests are run on the presence or absence of particular answers.

more authority reliant than other boys (Table 6, Item 5).

Finally, a crucial index of independence from the family is yielded when *S*s are asked to choose an adult ideal. As we would expect, the stable group, which includes all boys who choose the same occupation as the father's, also has the largest proportion of boys who choose the father as an ideal. The upwards more often choose a model outside the family, or describe a composite of characteristics from several people. More often than other boys, downward mobiles say there is no adult they wish to be like. And despite the fact that they choose the father less often than stable boys, they give the largest proportion of within-family models of any group, particularly choosing grandfathers and uncles. Here again we can observe the ambivalence of the downward mobile boys; though they are covertly rebellious toward family authority, they cling to within-family models. Another way of viewing this finding is to see in it a reflection of the downward mobile boys' narrowed and immature life space. Havighurst (4) has found that the tendency to choose within-family models characterizes the earlier years of adolescence. Our downward mobile boys retain the more infantile, more restricted image of the adult world.

The upward mobile boys, on the other hand, tend to a certain precocity in the loosening of family ties. Yet we have seen in the previous section on internalization that this is not associated with rebellion from or rejection of the family. Further evidence along these lines appears in the discussion of family milieu.

Until now evidence has been presented bearing on the *sense* of autonomy. Some other data suggest that these differences are related to the objective degree of autonomy permitted upward and downward boys by their parents. We asked: "What are your parents' ideas about the way you spend your money?" One fifth of the downward mobile respondents replied that they do not have funds of their own but ask their parents for money as they need or want some special thing. In the upward mobile group, fewer than one per cent gave this answer (Table 6, Item 1). Since there is no difference between the groups with respect to actual spending of money, we may infer that the difference resides in the *control* of spending.

Orientation Toward the Self

Upward mobile boys show a high degree of self-acceptance, and a confidence in social situations. The downward mobile is more ambivalent toward himself and more unsure and conflicted in social interaction.

Interviewers rated boys on a number of variables, among which were self-confidence, humour, and the clarity with which they organized and presented their opinions and attitudes. We may interpret these ratings, with caution, as reflecting the boy's social skill and ability to handle himself in social interaction with an adult. On each of the ratings, the upward mobile are high; they are more self-assured, show humour more often, and are better organized in the interview. The downward mobile are relatively unconfident, humorless, and disorganized (Table 7, Items 1, 2, and 3).

We find signs of self-rejection and demorali-

TABLE 7

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF ORIENTATION TOWARD THE SELF*

Item	Mobility			χ^{2*}	p Level
	Upward	Stable	Downward		
	(N = 277) %	(N = 168) %	(N = 73) %		
1. Self-confidence[a]				7.93	.10
a. high	30	23	18		
b. average	52	57	57		
c. low	18	20	25		
2. Humour				7.14	.05
a. present	74	67	58		
b. absent	26	33	42		
3. Organization of ideas[a]				7.31	.10
a. high	60	53	50		
b. average	22	27	20		
c. low	18	20	30		
4. Desired changes[b]					
a. self rejecting, major	04	10	15	6.36	.05
b. changes that are impossible	06	07	23	11.33	.01
c. changes within boy's own power	37	29	05	14.10	.01

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] In these cases, the chi square was derived from a 3 x 3 table with four degrees of freedom.

[b] Categories here are not exhaustive of responses given to an open question, and more than one response was often given. Tests are run on the presence or absence of particular answers.

zation in the downward mobile boys' answers to the question, "What would you like to change about yourself if you could—about your looks or your life or your personality?" They more often desire changes so gross or so central as to indicate alienation from the self; and they more often wish for changes that are unlikely to occur. The upward mobile boy more often refers to changes he has the power to effect himself. He is more realistically critical of himself, and less self-rejecting (Table 7, Item 4).

Family Milieu

The reader may recall that we felt far more confident of our predictions about the downward than the upward mobile in this area. We believed that the downward group would show a pattern of ambivalence towards the parent, a mingling of dependency and hostility; and we further held that this pattern would arise out of inconsistent, overly harsh methods of discipline. We did not, however, feel we could make any specific statement about the upward mobile: here we were interested in discovering what we could about the family milieu, in particular whether upward aspiration is associated with a rejection of family values, or a defensive identification with a forbidding parental authority, or by that relatively amiable relationship to the parent which permits the growth of autonomy.

TABLE 8

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF FAMILY MILIEU

Item	Mobility			χ^2*	p Level
	Upward	Stable	Down-ward		
	(N = 277) %	(N = 168) %	(N = 73) %		
1. Punishment[a]					
a. physical	02	08	15	19.12	.01
b. deprivational	66	69	65		
c. psychological	32	23	20		
2. Portrayal of parental figures					
a. harsh	27	35	48	11.59	.01
b. non-harsh	73	65	52		
3. Disagreements with parents					
a. have disagreements	66	56	49	6.39	.05
b. do not have disagreements	34	44	51		
4. Parents' attitudes toward way boy spends money					
a. disapproval	19	16	05	8.58	.05
b. no disapproval	81	84	95		
5. Leisure activities with parents[a]					
a. share many	19	11	09	17.23	.01
b. share some	73	74	68		
c. share none	08	15	23		

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] In these cases the chi square was derived from a 3 x 3 table with four degrees of freedom.

To begin with, we find that deprivation seems to be the dominant method of discipline for all groups; differences here are not significant. Striking differences appear, however, in the relative frequency of "psychological" and "corporal" methods of punishment. Only one in fifty of the upwards say that their parents use physical punishment; approximately one in ten of the stables and one in seven of the downwards report this method. In contrast, the use of "psychological" techniques (such as "given a good talking to") occurs in more than one quarter of the upward mobile responses, about one in five of the stables, and only one in eighteen of the downwards (Table 8, Item 1). We find that the parents of the upward aspirers avoid the harshest method of punishment; the "psychological" methods, which we suspect encourage the child's internal controls, are infrequently used by the parents of downward mobile boys.

Are our mobile *S*s presenting an over-idealized picture of their families? We looked for the pattern of repressed hostility and surface idealization which has been shown to characterize defensive identifiers (2, 5, 9). But our data reveal that the opposite is true: in the projective materials—where the *S*s are not discussing their own parents and where they have an opportunity to discharge repressed hostility without danger—the upward aspiring portray parents as less harsh than do the other groups (Table 8, Item 2). And as we shall see, they are more aware of differences between themselves and their parents. They more commonly report that their parents have some "old-fashioned ideas, or ideas they disagree with," (Table 8, Item 5), and that their parents in some respect disapprove of the way in which they spend their own money (Table 8, Item 6). We saw previously (Table 6, Item 3) that they are not likely to feel that family relationships are necessarily more intimate than friendships; and they less often choose an ideal adult from within the family group (Table 6, Item 6). The upward mobile boys are, then, more sensible of differences with parents.

Do the aspiring boys have a detached rather than congenial relationship to their parents? The single bit of evidence on this question suggests that this is not the case: they more often report engaging in leisure activities with their parents (Table 8, Item 5).

The downward mobile boy's family relationship is, as we expected, marked by ambivalence: the pattern is one of conscious idealization accompanied by unconscious suspicion and hostility.

We have already seen that the parents of this group are more often severe in their punishment. And in this group we find evidence of repressed hostility: in the projective stories, the downward mobile boys picture parental figures as both harsh and suspicious (Items 3 and 4, Table 8). Yet consciously they are more likely to deny differences with their parents. Some of our findings on autonomy have shown this; and in addition we find that these boys most often say they have no disagreements with their parents (Table 8, Item 5), and they least often say their parents disapprove of their handling of funds (Table 8, Item 6). The downward mobile boy shares leisure activities with his parents somewhat less than do boys in the other two groups (Table 8, Item 7).

Alternative Hypotheses

Before we discuss the possible meanings of these findings, we must consider briefly alternative hypotheses which might account for them. These are: (*a*) An age differential among the three groups; thus, the upward aspiring *Ss* might have given "little boy" job hopes, while the non-mobile group presented more mature aspirations; (*b*) a difference among the groups in frame of reference, so that the upward aspirers may have understood the interviewer to be asking for "fantasy" aspirations while other boys took the question to mean "realistic expectations"; (*c*) objective differences in the social status of the groups: perhaps the upward mobile boys' families are at the upper end of their group in income or education, and so provide their sons with an objectively greater opportunity for mobility; (*d*) a difference in intelligence among the groups.

First, on the question of realism. There were in our total sample fewer than one per cent of the boys who gave us their job choices distinctly "little boy" or "glamour" occupations. Among boys in the mobility analysis subsample, there were no aspirations listed which were obviously or totally based on ideas of glamour. When the boys are asked how sure they feel about the job they've chosen, we find that very few of them say they're not at all sure; more important, the variation that does occur is not related to the type of aspiration they express. Boys in the three groups show approximately equal certainty that they will *get* the jobs for which they are aiming (Table 9, Item 1).

Furthermore, we find a difference in educational plans, one which is appropriate for the difference in job aspirations; this indicates at least a minimal degree of realism in our *Ss*'

TABLE 9

RELATIONSHIP OF MOBILITY ASPIRATION TO AGE, CERTAINTY ABOUT ASPIRATION, FAMILY INCOME, FATHER'S EDUCATION AND INTELLIGENCE

Item	Mobility			χ^2*	p Level
	Upward	Stable	Downward		
	(N = 277) %	(N = 168) %	(N = 73) %		
1. Certainty about aspiration					
a. high	57	58	54	3.22	.70
b. moderate	34	33	33		
c. low	09	09	13		
2. Age					
a. fourteen	33	29	31	1.94	.80
b. fifteen	30	36	33		
c. sixteen	37	35	36		
3. Family economic status[a]					
a. high average	32	35	31	2.76	.70
b. average	53	49	52		
c. low	15	16	17		
5. Father's education					
a. grade school	32	27	25	3.07	.70
b. high school	50	58	55		
c. college	18	15	20		
5. Verbal ability					
a. high	25	11	03	33.92	.01
b. average	69	75	74		
c. low	06	14	23		

* Unless noted, the chi square was derived from a 3 x 3 table with four degrees of freedom.

[a] The measure used for economic status was the Remmers House and Home Scale.

understanding of work preparation (Table 3, Item 5).

The groups differ neither in age (Table 9, Item 2) nor in present socioeconomic status. The upward mobile boys are not from homes which are higher economically or in educational background (Table 9, Items 3, 4).

The two rough measures of intelligence available both show the striving group to be somewhat superior. The upward aspiring boy is more likely to report leisure reading (Table 2, Item 5), and he is rated by coders as more facile verbally (Table 9, Item 5). The downward mobile are rated low on verbal skills, but their reading pattern does not differ significantly from that of the stable group ($\chi^2 = .10$; $p > .90$ with 1 *df*). We will defer a consideration of this until the discussion.

DISCUSSION

We may interpret our data on the upward aspiring boy in relation to ego structure and functioning. We did not, of course, make a direct investigation of intrapsychic processes; furthermore, our understanding of "normal" ego functioning is, at present, limited and tentative. Nevertheless, it may be of some interest to attempt a formulation of our results in the current vocabulary of ego psychology.

Among upward aspiring boys we find a high level of diversified activity, suggesting that energy is available to the ego for focused use in work and play. We may infer that the antecedent condition for this accessibility is in the ego's having at its disposal a relatively high degree of neutralized energy. If we assume that both groups, mobile and non-mobile, do not differ in primary energy resources, we may conjecture that the aspiring *S*s have less energy tied to the resolution of conflict, and consequently gain in neutralized, disposable energy.[5]

Another sign of the comparative effectiveness of ego functioning among aspiring boys is seen in the greater degree of cohesion among the three major systems within the personality. On the one hand, ego processes in the mobile boy are less likely to be imperiled by the breakthrough of impulses (When would a boy break a rule?); on the other hand, he is less likely to show evidence of an excessively punitive super-ego. The greater degree of internalization of moral values among aspiring boys suggests a more harmonious articulation between the ego and super-ego systems. Our data show that this integration is not purchased at the expense of an unusual quantity of guilt.

A consideration of the intelligence factor is relevant here. Although we had no direct measure of intelligence, we saw that the upward mobile boys read more and read more difficult things; and they are judged on the basis of the interview to be more verbally skilled. In all likelihood there *is* an intellectual difference among our groups; one plausible interpretation of many of our findings is that they merely reflect group differences in IQ. We feel, however, that such a formulation assumes an overly simple causal sequence—for one thing it fails to clarify the mechanisms which mediate between IQ and other indices of successful ego functioning. A good deal of recent data points to a high degree of association between intelligence and variables which indicate effective ego functioning. The direction of causality and connection among these variables is by no means understood. Furthermore, to explain our data as a function of intelligence alone would not allow us to generate predictions in the area of family relations. If we assume, however, that our data are best understood by positing differences in intrapsychic organization—and particularly in ego structure and functioning—we may then posit specific expectations about family milieu.

We would expect—on the basis of the higher energy level and the more refined control of impulsivity among upward mobile *S*s—that they have relatively good relationships within the family and have successfully resolved infantile object ties. We mean by this that there is a relative absence both of covert hostility toward the parents, and of overt dependence upon them, together with a relatively high degree of objectivity about family relations.

We have seen that this is, in fact, the case. The mobile boy asserts (and probably has been encouraged toward) greater independence of judgment and behavior. It will be recalled that he expresses a desire for equalitarian rather than dependent relationships with adults. He frequently uses his own convictions as authority in making decisions.

---

[5] Cf. Rapaport (8, p. 353): "In other words, the amount of energy which the person can dispose by investing it in objects, by becoming interested in activities, even when essential drive-aims and drive-objects are in abeyance, is an indicator of ego-autonomy and ego strength."

The aspiring boy more often admits differences between himself and his family. He seems able to assert the legitimacy of an autonomous self-definition—to the extent, at least, of selecting an ego ideal outside the family setting.

The independence we have observed has not occurred through rebellion from or rejection of the family. Despite their ability to differentiate themselves from their parents, the upward aspirers give evidence of a congenial relationship with them. They share their leisure activities; in response to a projective measure they reveal relatively little covert hostility toward parental figures.

Their parents are, in fact, comparatively lenient; they are more likely to employ mild and essentially verbal discipline and use physical punishment infrequently.

We have seen, then, that the upward aspiring boy is characterized by a high energy level, the presence of autonomy, and a relatively advanced social maturity. These attributes may be viewed as derivatives of a generally effective ego organization, one which has developed out of the successful resolution of infantile object-ties and conflicts. We would expect, on theoretical grounds, that this development would be most likely to occur in a family setting where the parents allow the child a nonambivalent connection with them, where autonomy can emerge without conflict. As we have seen, the data we have suggest that this is the case.

We have already presented, in the introductory section, our view of the internal psychic processes which characterize downward mobile boys. From the specific perspective of ego functioning, we see an apparent blocking or impoverishment of energy which should, ideally, be available to the ego. We may infer that neutralization is not being successfully accomplished. There is a relatively poor articulation among the psychic systems: impulses threaten the ego's integrity; the superego seems overly severe and yet incompletely incorporated. These boys seem humourless, gauche, disorganized—relatively so, at least. Perhaps the most telling and poignant datum which the study locates is their response to the possibility of personal change, their tendency to want to change intractable aspects of the self, and the degree of alienation revealed by their desire to modify major and fundamental personal qualities. As we have seen, this pattern is associated with a certain quality of family interaction: the boy gives evidence of an ambivalent tie to the parents, a mixture of overt dependency and covert aggression. One of the determinants of this ambivalence, we may infer on theoretical grounds, is in the parents' relatively punitive style of discipline which fails to establish self-governing controls in the child.

A final point concerning the mechanisms through which the differing aspirations become established: In the case of upward aspiration we assume a direct connection between the socialization process and the aspiration to social mobility. That is, we assume that the parents transmit the value on mobility during socialization, and that the child incorporates this as part of a complex of values which includes autonomy and general achievement. In contrast, we conjecture that the process in downward aspiration may be more indirect: the rebellion engendered by an ambivalent family atmosphere may be directly expressed in a rejection of general social goals; or we may have a diffuse demoralization which causes (and prepares the child for) failure and isolation in social situations. In turn, these failures result in a general narrowing of goals, and a retreat from success; a central expression of this is the lowering of occupational goals.

SUMMARY

We have only recently begun to recognize that prevalent views of the dynamics of mobility, based as they are largely on literary sources and on extreme instances, require refinement and differentiation. The paper begins by distinguishing varieties of mobility behavior. There is presented a model of the psychodynamics of upward and downward aspiration, which is tested with data from a national sample of adolescent boys. The general formulation holds that upward mobility is found among boys with effective, autonomous ego functioning; downward mobility is seen as a symptom of demoralization. Psychoanalytic ego theory provides the framework

from which we make specific predictions of differences in the areas of activity level, achievement mode, time perspective, internalization of values, autonomy, self-esteem, and family milieu. The findings reveal sharp differences, in the predicted directions, between upward and downward aspiring groups.

REFERENCES

1. ACKERMAN, N. W., & JAHODA, MARIE. *Anti-Semitism and emotional disorder.* New York: Harper, 1950.
2. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D., & SANFORD, N. *The authoritarian personality.* New York: Harper, 1949.

.3 GOUGH, H. Personal communication, 1955.

.4 HAVIGHURST, R. J., ROBINSON, MYRA Z., & DORR, MILDRED. The development of the ideal self in childhood and adolescence. In J. Seidman (Ed.), *Readings in adolescence.* New York: Dryden Press, 1953.

5. HOFFMAN, M. L. Some psychodynamic factors in compulsive conformity. *J. abnorm. soc. Psychol.*, 1953, **48**, 383–393.
6. LYNN, K. S. *The dream of success.* Boston: Little, Brown, 1955.
7. McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
8. RAPAPORT, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1953.
9. SARNOFF, I. Identification with the aggressor: Some personality correlates of anti-Semitism among Jews. *J. Pers.*, 1952, **20**, 199–218.

Received March 4, 1957.

THE RELATION OF ACHIEVEMENT MOTIVATION TO PROBLEM-SOLVING EFFECTIVENESS[1]

ELIZABETH G. FRENCH AND FRANCIS H. THOMAS
AF Personnel and Training Research Center, Lackland Air Force Base

THIS study was designed to test the hypothesis that achievement motivation is a factor in problem-solving effectiveness. If motivation is assumed to have both energizing and directing powers, then an individual with high achievement motivation should both work harder and be more oriented toward identifying and utilizing cues relevant to the obtaining of an achievement-related goal than one without such motivation. A previous study by Thomas (2) demonstrated that achievement motivation was related to the length of time a subject would work on a problem without objective knowledge of progress but it did not show that it was related to reaching a successful solution. It was the purpose of the present study to test the hypothesis that achievement motivation is related to obtaining a problem solution as well as to the length of time spent on the problem.

METHOD

Measure of Motivation

The measure of achievement motivation was a projective device called The Test of Insight (1) in which the *S* is asked to provide "explanations" for briefly described behaviors presented as typical of hypothetical individuals The number of achievement-relevant explanations make up the score.

The Problem

The problem-solving apparatus consisted of a panel containing seven lights arranged in a horizontal row with a switch below each light. The wiring of the panel was such that some switches operated one light; some, two; and some, three. Similarly, some lights were controlled by one switch only and some by two or three. Operation of a switch always changed all the lights to which it was connected. Thus, if Light 1, controlled by Switch 1, was on and Light 3, also controlled by Switch 1, was off, operation of Switch 1 would reverse the pattern. *S*'s problem was to create a specified pattern of lights.

Subjects

The *S*s were 100 airmen undergoing basic training at Lackland Air Force Base. All *S*s selected had Armed Forces Qualification Test (AFQT) scores in the top 11 per cent of the airman population at large, in order to minimize the effect of intelligence on problem solution.

Procedure

Each day's *S*s, about 10 in number, were first seen as a group and given an orientation to the problem-solving that clearly indicated that good performance could be considered relevant to achievement goals. Then they took the test of Insight, described to them as unrelated to the main part of the experiment.

Upon completion of the test of Insight, each *S* was taken into the problem-solving situation individually. *E* demonstrated the apparatus to *S* and explained the important points, i.e., that switches might operate more than one light and vice versa, and that operation of a switch always changed all the lights it controlled. Each *S* was then given a practice period in which he was instructed to push the switches in order, one at a time, until he could demonstrate by reciting the combinations to *E* that he knew which switches controlled which lights. The time required for this varied from two to 12 minutes with a mean of 5.72; and the number of trials, from two to 11 with a mean of 4.98.

When *S* had demonstrated that he was in possession of all the necessary information to solve the problem, he was given the following instructions:

> Now I am going to give you the problem which will be to indicate which switches must be in the "up" position in order to produce a particular pattern of lights. But first I want to explain how you must do this. In working out the solution you may operate the switches if you like but you may have only one of them in the up position at a time. In other words, you may not work out the complete solution on the board. You must figure it out in your head from what you know of each switch and its lights. When you think you have the answer, write the numbers of the correct switches on a sheet from this tablet and drop it in the box. Write only the answer; you may not figure on the paper. You will have 35 minutes in which to solve the problem. You may use as much of that time as you like and you may submit as many answers as you like in that time. Put each answer on a separate sheet and drop each in the box before continuing. Please number the sheets in order.
>
> Is the procedure clear? You are to list the numbers of the switches which must be in the 'up' position to give the pattern I will ask for; you may submit as many answers as you like in the 35-minute time limit; and if you operate the switches in working out your solution you may have only one in the 'up' position at a time. If there are any questions ask them now because I

---

[1] This report is based on work done under ARDC Project No. 7719, Task No. 17011, in support of the Research and Development Program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

can't answer them after I have given you the problem.

The problem is, "What switches must be in the 'up' position to have the greatest possible number of lights lighted?"

E then started a timer and *S* began to work. *E* recorded the time and number of each solution as it was dropped into the box. *E* did not look at the solutions given and *S* had no external indication as to whether he was right or wrong. Those *S*s who had not stopped voluntarily by 34 minutes were given a warning that the time would be up in one minute in order to permit them to write down a solution if they cared to.

The actual solution required the operation of five of the switches, which lighted all lights. The solution could not be reached with either of the other two switches in the "up" position. The solution also involved turning off lights already on and getting them on again with another switch. The switches could be operated in any order.

In summary, *S*s in possession of all the information necessary to solve the problem and of approximately equal intellectual ability but varying in level of achievement motivation were given up to 35 minutes to solve a problem presented to them as a measure of competence.

RESULTS

The analysis was based on the data from 92 *S*s. Two were discarded because their records indicated they had failed to follow instructions. The records of six others suggested that they had given a correct solution by chance; i.e., a correct solution was given in the middle of a long series of incorrect ones and was not repeated. Since these *S*s could be classed as neither solvers nor nonsolvers, they also were eliminated. An *S* was considered a solver if he submitted one correct solution and stopped work or if he continued to work after his first solution, submitting repetitions or alternate versions. Since the five necessary switches could be operated in any order, any one of the possible orders was correct. Some *S*s gave several of them. Of the 92 *S*s remaining, 39 solved the problem and 53 failed to solve it. Reaching the solution was unrelated to practice time (r_{bis} .05), number of practice trials (r_{bis} .05), AFQT (r_{bis} .18), time spent on the problem (r_{bis} .13), using all the time (chi square .01) or number of solutions given (r_{bis} .02).

TABLE 1

NUMBER SOLVING IN HIGH AND LOW MOTIVATION GROUPS AND MEAN MOTIVATION SCORES FOR SOLVERS AND NONSOLVERS

Attainment of Solution	Achievement Motivation		Motivation Scores	
	High	Low	Mean	*SD*
Solvers	25	14	9.28	3.95
Nonsolvers	22	31	7.34	3.62
	χ^2 = 4.45		t = 2.31	
	p = .05		p = .03	

TABLE 2

MEAN TIME WORKED AND NUMBER STOPPING BEFORE THE TIME LIMIT FOR HIGH AND LOW MOTIVATION GROUPS

Achievement Motivation	Time Worked		Stopped before Limit	Worked until Limit
	Mean	SD		
High	27.14	10.02	25	22
Low	13.79	8.13	44	1
	t = 6.99		χ^2 = 23.22	
	p = .001		p = .001	

To test the main hypothesis, the *S*s were divided into groups with high and low achievement motivation by dichotomizing the distribution of motivation scores at the median. The 47 *S*s with scores of eight or above were called the high motivation group; the 45 with seven or below, the low motivation group. The proportion of solvers and nonsolvers in each motivation group is given in Table 1. Although approximately the same number of *S*s solved and failed in the high motivation group, less than a third of the low motivation group solved. The chi square for this analysis is significant beyond the .05 level. Table 1 also gives the achievement motivation means for the solvers and nonsolvers. The difference between them is significant at the .03 level. The relation between achievement motivation and attainment or nonattainment of a solution may also be expressed by a biserial correlation, which proved to be .30. These data were considered to support the hypothesis that individuals with high achievement motivation are more likely to succeed in solving a problem that is related to achievement goals than are individuals with low achievement motivation.

Since the *S*s were permitted but not required to work for 35 minutes, and the actual time worked ranged from three to 35 minutes, it was possible to check the earlier finding that motivation was related to length of time the *S*s worked. The data for the relation between motivation and time worked are presented in Table 2. The mean time worked for the high motivation group was 27.14 minutes as compared to 13.79 minutes for the low motivation group. The correlation between motivation score and time worked was .41. When time worked is considered in terms of whether the *S* stopped before the time limit or continued until stopped by *E*, the results are even more striking. Only one of the 45 low motivation *S*s worked until stopped, as compared with 22 of the high motivation *S*s, yielding a chi-square value of 23.22. The biserial correlation between motivation score and stopping versus not stopping was .54. Data for number of solutions offered by high and low motivation groups were essentially the same.

This relation between achievement motivation and time worked does not mean that the relation between motivation and problem success is solely a function of the fact that the more highly motivated *S*s worked long enough to arrive at a solution. In the first place, among those who did reach a solution, time spent on the problem was not necessarily the same as the time to the problem solution. Some *S*s continued to work beyond the first solution as described earlier. Thus, the mean time worked for the high motivation solvers was 25.2 minutes as compared to 18.7 for the low motivation solvers, a difference significant at the .05 level. However, the mean *solution times* for the two groups were 19.7 and 18.4 minutes, respectively, a difference that is not significant. In approximately the same period of time, 25 high motivation *S*s as opposed to 14 low reached the solution. Also, as pointed out earlier, there was no overall relation within the time limits of the experiment between the length of time worked and successful solution.

Nevertheless, the fact that highly motivated *S*s work longer obviously does have general implications for prediction of success. Consider, for example, the performance of the nonsolvers in the two motivation groups. All of the 31 nonsolvers in the low motivation group stopped before the time limit was up. Since they had stopped, there was no chance that they would reach a solution. Of the 22 nonsolvers in the high motivation group, 13 were still working at 35 minutes. Since the distribution of solution times ranged up to 35 minutes, it is reasonable to suppose that some of these nonsolvers would have reached a solution if permitted to work longer.

Three other differences between the high and low motivation groups were quite clearly suggested although they were not specifically predicted. The first is relevant to the point just discussed. Although there was no relation between time worked and reaching a solution (r_{bis} = .13), for the group as a whole the biserial correlation for the high motivation group only was −.28 and that for the low, .53 (Table 3). The former is significant at only about the .10 level, but the latter is significant at beyond the .01 level, and the difference between them is large enough to be suggestive of a real difference. The majority of the high motivation group but only a few of the low motivation group used most or all of the time, and in each group approximately equal numbers of the long working *S*s solved and failed to solve the problem. The big difference between the high and low motivation groups occurred among the *S*s who used only part of the time. Of those high motivation *S*s who worked a relatively short time, nearly all solved the problem. Only a small portion of the low motivation *S*s, who worked for a short time, solved.

TABLE 3

BISERIAL CORRELATIONS BETWEEN SOLUTION AND TIME, AFQT AND MOTIVATION FOR THE MOTIVATION GROUPS SEPARATELY

rSolution vs:	High Achievement Motivation	Low Achievement Motivation
Time	−.28	.53
AFQT	.36	−.08
Motivation	.04	.41

A second difference between the groups not specifically predicted was the relation between AFQT and problem success. The AFQT range had been kept small so that performance would not be affected by differences in intelligence and as stated, the relationship is represented by a nonsignificant biserial correlation of .18. For the high motivation group only, even with the restricted AFQT range, the correlation was .36 (p = .05). The correlation for the low group was −.08 (Table 3). The difference here is again large enough to indicate a point for further investigation. The suggested interpretation is that the *S*s in the highly motivated group are all trying hard so that even small differences in ability become important. On the other hand, in the low motivation group ability differences are not important because the *S*s are not using their potential as effectively as they could. This finding needs to be checked over a wider ability range before much can be said with assurance.

An additional finding which corroborates the notion that the high motivation group is more or less uniformly motivated, regardless of difference in scores, is a correlation of .04 between motivation score and solution for the high group. The low group had a correlation of .41 between motivation and solution (Table 3). About half the *S*s at each interval above the median solved the problem, while the proportion solving decreased with each interval below the median. Apparently, for this particular task, which had some intrinsic interest for most *S*s, achievement motivation of only about median intensity was necessary for optimum performance. A less interesting task might require a higher level or a more interesting one, a lower.

These last findings can be considered merely tentative, but they do point the way to fruitful areas for further investigation.

SUMMARY AND CONCLUSIONS

The hypothesis that when a problem is related to achievement goals, achievement motivation is related to problem-solving effectiveness and to the length of time *S*s will work on a problem without knowledge of results was tested. *S*s with high and low achievement motivation scores worked on a problem after first learning all the separate facts necessary for solution. They were permitted up to 35 minutes.

The results showed that *S*s with high achievement motivation were more likely to reach a solution than those with low motivation. High motivation *S*s also worked longer than low motivation *S*s. The relation between motivation and success was shown not to be a result of the relation of motivation and time worked since actual solution time for those who solved was not different for the two motivation groups. Other findings not specifically predicted are discussed and tentative interpretations are offered.

REFERENCES

1. FRENCH, ELIZABETH G. Development of a measure of complex motivation. *Res. Rep. 56–48*, AF Personnel and Train. Res. Center, April 1956, Lackland Air Force Base, Tex.
2. THOMAS, F. H. Visualization, experience, and motivation as related to feedback in problem solving. *Amer. Psychologist*, 1956, **11**, 444. (Abstract)

Received March 25, 1957.

THE EFFECT OF BRIEF SOCIAL DEPRIVATION ON BEHAVIORS FOR A SOCIAL REINFORCER

JACOB L. GEWIRTZ[1]
National Institute of Mental Health
AND DONALD M. BAER[2]
University of Washington

DOES the incidence of behavior for social contact with other people increase following a period of social deprivation? Stated in more general terms, are there social drives that function in a manner similar to the primary appetitive drives? This is the central question to which this investigation addresses itself, employing young children as subjects (*S*s).

Short-term procedures (other than learning operations) that account for variance in the effectiveness of a reinforcer (or in behavior output) are ordinarily labeled *drive* operations. One notable operation of this type in the case of the primary appetitive drives evolves from the fact that relevant stimuli (e.g., food, water) acquire maximal reinforcing value for an organism only subsequent to its recent deprivation of them. Other reinforcers of no apparent biological importance, like that provided by the opportunity to make a brief observation response, appear also to be raised in effectiveness following preceding periods of deprivation (2). In this context, it is a provocative question whether *social* reinforcers (those dispensed by people), postulated to possess reinforcing value through a history of conditioning, respond in a similar manner to deprivation. The relevance of this hypothesis accrues both from these precedents and also from the extensive use of the deprivation concept in a number of conceptual systems dealing with general behavior (e.g., 11, 13, 18, 19) and social behavior (e.g., 1, 7, 17, 20).

There is a considerable body of evidence (3, 8, 12, 21, 22) showing that an experimenter can increase the relative frequency of an arbitrarily chosen response of a child by dispensing certain social stimuli (words and phrases) contingent upon that response. In particular, the social reinforcer[3] *approval* is readily dispensed through the use of simple words and phrases like "Good!" "Fine!" and "Mm-hmm." Furthermore, approval appears to be representative of the reinforcers that control the purely social initiations made by children to adults.

For these reasons, this study is concerned with the possible responsiveness of the social reinforcer *approval* to a condition of social isolation, representing the deprivation of all social reinforcers (including approval).

METHOD

Each *S* served twice, once under a condition of Deprivation and once under a condition of Nondeprivation. Four discontinuous independent variables representing levels of Deprivation, the Sequence of the Deprivation and Nondeprivation conditions, the Sex of *E*, and the Sex of *S*, generated a 2 x 2 x 2 x 2 *factorial* design. A score reflecting the effectiveness of approval as a reinforcer was the dependent variable.

Four continuous independent variables, representing *S*s' characteristic Attention, Approval, and Achievement orientations, and their Ages, were related to *changes* in this same dependent variable caused by Deprivation, in a *regression* analysis. (Each *S* having served under both Deprivation and Nondeprivation conditions, the *change* in the effectiveness of approval assignable to Deprivation can be taken as the difference between the scores gathered from each *S* under these two conditions.) The detailed descriptions of these variables follow.

FACTORIAL DESIGN

Independent Treatment Variables

Social deprivation. On one of the two occasions when each *S* came to play, he was subjected to a 20-minute period of social isolation before the game (*Deprivation*).

---

[1] This study, which was carried out when the senior author was at the University of Chicago, was facilitated by a grant-in-aid to him from the Social Science Research Committee of that institution. The authors are grateful to Mrs. Chaya Roth for her very helpful assistance.

[2] At the time of this study the junior author was a Public Health Service Predoctoral Research Fellow of the National Institute of Mental Health at the University of Chicago.

[3] According to the logic used here, many apparent complexities of social contact and dependence may be referred to the assumption that stimulus events provided by the positive social responses of people (e.g., attention, approval, affection) function essentially as positive *reinforcers*. That is, the supply of these stimuli contingent upon various behaviors can systematically increase the outputs of those behaviors (6).

On this occasion, when S (escorted by E) arrived at the experimental room, a familiar adult would appear to announce that the toy was "broken" and was "being fixed." S was given plausible assurance that the toy would be repaired "in a few minutes," and that since E wanted him not to miss his turn with the game, he might as well wait in the experimental room, since the adjoining "repair" room was clearly crowded. S was then left to wait alone without playthings in the experimental room with the door closed (or slightly ajar if he was more comfortable that way), while both adults retired to "repair" the toy. (During this process, occasional appropriate noises were emitted to indicate that repairs were progressing.) This procedure was carefully structured so as not to imply sanctions, and appeared adequate to guard against the possibility that S might feel rejected. If S poked his head out and asked to leave during the 20-minute isolation period, he was warmly reassured that the toy was "almost fixed," and he returned to the room. (This rarely occurred.) After 20 minutes, E returned with the "repaired" toy, and the game was played in the usual fashion. Under the control condition (*Nondeprivation*), S began the game immediately upon his arrival at the experimental room from his nursery group.

Sequence. Each S played three games with a given E, the last two separated by at least one day. The first of these, a Nondeprivation condition, was to introduce S to the game and to E, and to give him some familiarity with the experimental setting. (Data from this adaptation session are not included in the analysis.) For half of the Ss, the second session was preceded by the social isolation condition, and the third was not; for the other half, this order was reversed.

Sex of experimenter. Half of the Ss in the sample played their three games with a male E, and the other half with a female E. Both Es were graduate students in their early twenties. In dress and appearance, they were not unlike typical adults in the nursery school environment.

Sex of child. Each E served with four boys and four girls under each sequence.

Dependent Variable

After the experimental treatment of Deprivation or Nondeprivation, the task for S was to drop marbles into a toy while E looked on. The toy was a wooden, angular box, having two holes at its forward top into which marbles could be dropped, and an open tray at its base where they returned. The tray contained six marbles, identical in size but different in color, which could be used repeatedly. The experimental room contained two chairs, a couch, and a low table upon which the toy rested. Each S was introduced to this setting by E as if to a game. E and S sat before the toy, and S was told that he could drop marbles down either hole, but that he must drop only one at a time. E then observed S's play for a "baseline" period, during which reinforcers were *not* dispensed, until S's responses appeared relatively stable (typically by four minutes). Meanwhile, E responded to S's comments and questions in a friendly but brief manner. Without pause, the 10-minute test of reinforcer effectiveness began, as a continuation of the baseline period. E proceeded to dispense the reinforcer according to a schedule immediately after S dropped marbles into the hole preferred *least* during the last minute of the baseline period (the *correct* response).

Reinforcement. The reinforcer, designed to appeal primarily to the concept of approval, consisted most frequently of the word "Good!" and less frequently of such phrases as "Hm-hmm," "Good one!" "That's a good one!" and "Fine!" These reinforcers were delivered by E in a casual manner, according to a schedule incorporating four successive fixed-ratios: 1. At the beginning of the first minute of reinforced play, every correct response was reinforced until 10 consecutive correct responses had been made. 2. Subsequently, five reinforcers were delivered, one for every *second* correct response. 3. Then, five reinforcers were delivered, one for every *third* correct response. 4. Finally, a reinforcer was delivered for every *fifth* correct response until the end of the 10 minutes of reinforced play. This schedule was qualified by three conventions: 1. Every correct response that immediately followed an incorrect response was reinforced. 2. If S made five consecutive incorrect responses, E began the schedule again, halving the number of reinforcers called for at each ratio level, until the reinforcement ratio last employed was again in effect. 3. If S made 20 successive incorrect responses, E said, "You can put the marbles in *either* hole."

Reinforcer effectiveness score. The numbers of marbles dropped in the correct and incorrect holes during each minute of play were recorded during the game (which immediately followed the deprivation or control treatment) by an observer in another room. These basic data generated from one to five pairs of scores for the baseline level period, and 10 pairs of scores for the reinforcer effectiveness test period. For each minute of play, the *relative frequency* of a correct response was the ratio of correct responses to total (i.e., correct + incorrect) responses. The *median* of these 10 test ratios was employed as a single representative score reflecting reinforcer effectiveness *uncorrected* for baseline level. The baseline level was taken to be the ratio from the final minute of the baseline period, when no reinforcers were dispensed. (This ratio was assumed to be maximally free from possible "settling down" variations expected during the initial minutes of play, and hence to display the greatest continuity with the following 10-minute reinforcer period.) The corrected effectiveness of the social reinforcer was assessed by taking the difference between the median of the 10 reinforcer test ratios and the baseline ratio. This *reinforcer effectiveness score* represents the gain in relative frequency of the correct response attributable to the social reinforcer of approval.

REGRESSION DESIGN

Independent Variables Reflecting Individual Differences

Five experienced nursery teachers were asked to rate each of the Ss in their groups for the degree to which S characteristically displayed behaviors apparently aimed at securing attention or approval, or indicating a high achievement standard. (These characteristics were thought to represent possible determinants of the reinforcing effectiveness of approval for Ss.) It was not practicable to ascertain the reliabilities of the ratings. The instructions to the raters stated that the three scales were conceptually independent. All children in each class were rated on each scale at one time; ratings were made on continuous 100 cm. lines subdivided for reference into five areas. These ranged from "completely indifferent" through "slight," "moderate," "considerable," and "extreme" orientations. Each teacher's ratings were then transformed into a set of standard scores, for each scale separately. (These standard scores were constructed from ratings of *all* the children in the nursery school, of whom the experimental *S*s represent a sample.) The descriptions of these dimensions for the rater were as follows:

Seeks attention or notice from teacher: The child seeks attention or notice from the teacher, in settings where such attention or notice does not imply approval; e.g., "Watch me open this book," and "Look at me" (when the child is not doing something for which he could be praised). The child may deviate, cry, or whine to get the teacher to attend to him.

Seeks approval from the teacher: The child seeks to gain approval or recognition (praise) from the teacher either by direct request (e.g., "Am I not a nice boy?") or by performing in a praiseworthy manner (good behavior, achievement, and the like).

Has high achievement standards: The child who has set high standards for himself works for the sake of excelling at tasks (for which he may or may not receive recognition), is autonomously involved in his achievements, and would attempt new tasks and problems for the chance of solving them successfully.

Age. The age of each *S* was recorded, yielding a set of scores ranging from 46 to 63 months.

Dependent Variable

Deprivation score. The *increase* in the reinforcer effectiveness score from the Nondeprivation condition to the Deprivation condition serves as the dependent variable in the regression analysis. It is computed for each *S* by subtracting the reinforcer effectiveness score displayed under Nondeprivation from that displayed under Deprivation. This *Deprivation score* reflects for each *S* the gain in relative frequency of the reinforced response that may be attributed to the condition of social deprivation.

SAMPLE

Thirty-two middle-class *Ss*, ranging in age from 3-10 to 5-3, were selected at random from play groups of a university nursery school. Two of the 16 girls were Negro, the rest of the girls and all 16 boys were white. Otherwise, the primary criterion used in selecting any *S* was his willingness to participate in the "game" on the three occasions comprising the experiment.

RESULTS

Approval as a Reinforcer

A necessary preliminary question is whether the reinforcer dispensed was effective as such, independent of the Deprivation condition. The eight mean reinforcer effectiveness scores under the Nondeprivation conditions were tested in a 2 x 2 x 2 factorial analysis of variance (Sequence × Sex *E* × Sex *S*), and could be considered as drawn from the same population. The hypothesis then was tested that the mean reinforcer effectiveness score under the combined Nondeprivation conditions was zero, employing an error term based upon those scores and a one-tail t test. That mean, .10, was found to be reliably greater than zero at $p < .025$, supporting an affirmative conclusion.

TABLE 1

REINFORCER EFFECTIVENESS SCORE MEANS

Sex of *E* and *S*	*N*	Condition	
		Nondeprivation	Deprivation
Male *E*			
Boys	8	.13	.06
Girls	8	.17	.25
Boys & Girls	16	.15	.16
Female *E*			
Boys	8	−.04	.34
Girls	8	.15	.18
Boys & Girls	16	.06	.26
All children	32	.10	.21

Note:—Each *S* contributes both to Nondeprivation and Deprivation means.

Factorial Analysis

The means of the reinforcer effectiveness scores for the experimental conditions of Deprivation, Sex of *E*, and Sex of *S*, the conditions which produced reliable effects, are presented in Table 1. The Sequence variable did not enter into any reliable relationships, and was collapsed for this report. Each *S* yielded two scores, one under Nondeprivation and one under Deprivation. Table 2 presents a summary of the analysis of variance that provides the basis for evaluating differences among the group means. The statistical model employed in the analysis was an extension of Lindquist's Type III (14, p. 281).

The Deprivation variable contributed reliably to the variance in reinforcer effectiveness scores at $p < .025$. The mean reinforcer effectiveness score under Deprivation (.21) was twice that under Nondeprivation (.10). However, the effect under Deprivation varied as a function of which *E* delivered the reinforcer, according to the reliable first-order interaction effect between the variables Deprivation and

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF REINFORCER EFFECTIVENESS SCORES

Source of Variation	df	Mean square
Between subjects	31	
Sequence (Seq)	1	1,234
Sex S	1	736
Sex E	1	8
Seq × Sex S	1	284
Seq × Sex E	1	2
Sex S × Sex E	1	395
Seq × Sex S × Sex E	1	1,324
Residual	24	1,531
Within subjects	32	
Deprivation (Dep)	1	1,795**
Dep × Seq	1	11
Dep × Sex S	1	479
Dep × Sex E	1	1,671*
Dep × Seq × Sex S	1	570
Dep × Seq × Sex E	1	820
Dep × Sex S × Sex E	1	2,438***
Dep × Seq × Sex S × Sex E	1	62
Residual	24	297

* Yields F ratio significant at $p < .05$.

** Yields F ratio significant at $p < .025$.

*** Yields F ratio significant at $p < .01$.

Sex of E ($p < .05$). A reliable second-order interaction effect at $p < .01$ provides a final qualification by sex membership of the child and adult on the Deprivation effect, and clarifies the nature of the Deprivation × Sex of E interaction effect. Social deprivation enhances the effectiveness of approval as a reinforcer essentially as a cross-sex function: the reinforcer dispensed by each E appears on the average more effective after Deprivation only with Ss of the opposite sex, with this difference reliable solely for boys with the female E. With Ss of the same sex, the effect of Deprivation on the reinforcer appears negligible (and, in the case of boys with the male E, apparently in the reverse direction). Appropriate one-tail t tests indicate that the reinforcing effectiveness of the male E with boys under Deprivation is questionable ($p < .15$); in all other conditions, the male E has clear reinforcing value at $p < .05$. The female E appears to have no reinforcing value for boys under Nondeprivation, but is reliably increased by Deprivation ($p < .01$) to greater than zero effectiveness ($p < .05$). With girls, the reinforcing value of the female E is questionable ($p < .15$ under Nondeprivation, $p < .10$ under Deprivation).

Regression Analysis

It will be recalled that the *Deprivation score* reflects for each S the increase in reinforcer effectiveness attributable to the condition of Deprivation. The four continuous independent variables of Age and Attention, Approval, and Achievement orientations were entered into a multiple regression analysis, with the Deprivation score as the dependent variable.[4] (This analysis requires the usual assumptions about the simple regressions involved (9, p. 170), as well as the assumption of homogeneous linear regressions for the eight cells of the 2 x 2 x 2 factorial in the case of every pair of those five variables.) The analysis demonstrates the independent contribution of each of these variables to the Deprivation score, and for the present problem focused only on the identity of the partial regression coefficients found to be reliably different from zero. The Achievement variable was eliminated because its partial regression coefficient was not reliably different from zero. The analysis then yielded a corrected (9, pp. 433–34) multiple correlation coefficient of .82 (20 df, $p < .001$).[5] The partial regression coefficients of Age, Approval, and Attention are reliably distinct from zero (at $p < .001$, $p < .01$, and $p < .01$ respectively, for 20 df). (These were evaluated by calculating t as the ratio of each coefficient to its standard error.) The regression equation was:

Deprivation score (predicted) = 3 (Age) + 1 (Approval) − 1 (Attention) − K. (The

---

[4] This analysis employed the residual moments (variances and covariances) of these scores, which resulted from isolation of the moments generated by the classifications of the 2 x 2 x 2 factorial design (Sequence, Sex of S and Sex of E), with the Deprivation and Nondeprivation conditions collapsed to form the Deprivation score. Thus the analysis is essentially independent of the levels determined by those experimental classifications, and represents a distinct further analysis of social deprivation, reflected now in the dependent variable.

[5] Degrees of freedom for the multiple correlation coefficient and for each partial regression coefficient were computed as the total number of Ss (32) less the number of effects generated by the factorial classifications of the Deprivation score (seven: three main effects, three first-order interactions, one second-order interaction) less the number of variables involved in the multiple regression (five including Achievement, four excluding it) less one.

TABLE 3

RESIDUAL INTERCORRELATIONS OF AGE, ATTENTION, APPROVAL, ACHIEVEMENT, AND DEPRIVATION SCORES[a]

	Attention	Approval	Achievement	Deprivation
Age	.44*	.49*	.32	.75***
Attention		.36	.41*	.11
Approval			.34	.58**
Achievement				.20

[a] Each r, determined from the residual moments of analyses of covariance, has 23 df. Significance levels are denoted as follows: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. An r of .34 is reliable at $p < .10$.

final term of this equation is a constant specific to the data of this experimental case.)

The 10 simple (residual) intercorrelations among the four independent variables and Deprivation are presented in Table 3. The correlations of the Deprivation score with Age (+.75, $p < .001$) and Approval (+.58, $p < .01$) are seen to be reliable and positive, while the correlation of Deprivation with Attention (+ .11) is not reliably different from zero. In addition, the low correlations of Attention with Age (+.44, $p < .05$) and with Approval (+.36, $p < .08$) are positive and similar in magnitude and reliability. Together with the reliable negative partial regression coefficient found for the Attention variable ($p < .01$) this correlation pattern can only indicate that Attention functions as a *suppressor* variable (15), to subtract from the variables Age and Approval a portion of the variance in each which is not related to the dependent variable of Deprivation. In this way, the regression coefficients of Age and Approval on Deprivation would be enhanced, the former somewhat more than the latter. Hence, the negative partial regression coefficient for Attention does not represent a negative relationship to Deprivation.

DISCUSSION

As in previous studies, the reinforcing character of verbal approval has been demonstrated. It has been assumed here that approval had acquired reinforcing value through *S*'s previous history of conditioning. It might be expected therefore (*a*) that approval would function as a more effective reinforcer than would (for example) a comparable noise emitted by a machine, and (*b*) that such a noise would be less affected by a preceding period of social isolation. (A demonstration of both of these points clearly would strengthen the conclusions to be drawn in this study.) Similarly, verbal approval from adults has been assumed to represent the social reinforcers controlling the purely social initiations made by children to adults, and the generality of the conclusions must necessarily rest on this assumption.

The effect of social deprivation upon approval as a reinforcer. The data appear to support the conclusion that the effectiveness of a social reinforcer may be increased by the deprivation of social reinforcers generally. This conclusion is immediately qualified by the observation that it holds differentially according to Sex of *E* and Sex of *S*. It is clear that the conditions under which the effect obtains remain further to be delimited, since sex membership in itself hardly explains the matter.

Further, it is important to attend to the postulated equivalence of social isolation and the deprivation of social reinforcers. Clearly, isolation implies the absence of all agents of social reinforcement. Even so, it would be difficult to deny that isolation may imply other characteristics significant to the behavior observed here. For example, isolation might be expected to increase the vigilance of the child, to set up a condition of frustration or annoyance, or to give an impression of reproof or disapproval, and any of these concepts might yield predictions in accord with the data. However, other aspects of the data at hand seem to support the equation of isolation to the deprivation of social reinforcers.

One such aspect is afforded by the (partial) regression of the children's approval and attention orientations on the Deprivation score, which reflects the effect of social deprivation on the reinforcing power of approval in the game. It was found that social deprivation increased reliably the reinforcing power of adult approval for *S*s as a positive function of the degree to which they typically sought approval from the nursery teacher (and presumably from similar adults).

Let us assume for the moment (*a*) that social deprivation represents the unavailability of both approval and attention (among other social reinforcers), and (*b*) that those reinforcers are not equal alternatives for *S*, i.e., that there is for each *S* a reinforcer hierarchy such that some one reinforcer has primary appeal for him. Then those *S*s who typically seek approval would have been most susceptible to the reinforcer supplied in the game after isolation; but those *S*s who typically seek attention may have found that *E*, through his constant attentive interest, was dispensing a uniform amount of attention independent of their behavior in the game. Since the reinforcer offered contingent upon a correct response would be essentially irrelevant to these attention-seeking *S*s, then no gain in control of that response in the game would follow from the condition of social deprivation. Rather, the *S*s whose attention-seeking tendencies increased as a function of isolation may have employed other behaviors aimed at securing more attention from *E* than the standard amount offered in the experimental setting. And these behaviors, unscored in this study, would have competed with the game-playing response for approval. On this basis, it would be expected for this type of experimental setting that the approval-seeking tendencies of *S*s would relate positively, and their attention-seeking tendencies would relate negatively, or not at all, to the effects of social deprivation upon the reinforcing value of approval. This pattern is in accordance with the conclusions yielded by the multiple regression analysis.

In other words, the results are consistent with the premise that social deprivation sensitizes primarily those social stimuli which are in fact reinforcers for the child deprived. This reasoning could lead to a more precise statement of the experimental conclusion: the effectiveness of a social reinforcer may be increased by its own deprivation. This suggestion might be checked in a design that deprived *S*s of particular social reinforcers (e.g., the one most important for them) but not of all others. If replicated in a sufficient variety of settings, the suggestion could perhaps provide a criterion for a conditioned drive convention, insofar as it specifies an operation of deprivation for a presumably conditioned reinforcer paralleling the criterion of a standard drive operation of deprivation for primary reinforcers. The validity of such a convention would be increased by demonstrating the responsiveness of specific social reinforcers to the *satiation* operations against which the effect of deprivation typically would be reflected; this is the object of a companion study.[6]

Furthermore, a similar effect was found in a previous study (5) of presumably related phenomena involving different *E*s from those of this study and *S*s in the same age range. In that experiment, individual children engaged in easel painting could emit attention-seeking responses to an adult under conditions of the apparent low and high availability of the adult as a source of reinforcers. Similar control was exercised over the sex membership of the *S*s and the two adults of each sex. The same cross-sex pattern of behavior for a social reinforcer class (in that case, attention) was observed, the low availability condition effecting an increase in these behaviors relative to high availability in a manner very similar to the effect of the Deprivation condition in this study. Hence, it appears reasonable to emphasize what appears basic to both conditions, the relative availability of, or access to, social reinforcers. On this basis, social deprivation could be conceived as an operation that need not require isolation for its implementation. That is, the relative availability of social reinforcers might be manipulated in a variety of ways, some involving the actual presence of an agent of reinforcement, as in the low adult availability condition of (5).

Finally, it may be important to consider the possibility that some *S*s might have accidentally experienced mild reproof, rejection, boredom, or the like at *E*'s hands. If such a class of variables were to increase sensitivity to social stimuli like approval or attention (cf., e.g., 10, 16), then it would show an increased reinforcing power, but not because of the operation of deprivation (or of *E*'s low availability) as such. Thoughtful care was taken to guard against these possibilities. The setting was designed to be natural and permissive, and *S*s were handled with warmth and reassurance. The possibility that some *S*s might have felt reproved, bored, rejected, etc. cannot be ruled out, but the generally good-natured responses of the *S*s to isolation make these effects seem unlikely.

---

[6] In the companion study to be published entitled, "The Effects of Deprivation and Satiation on Behaviors for Social Reinforcers," each of three independent groups of 34 *S*s, aged 6-6 to 9-0, served with a female *E* under similar Deprivation and Nondeprivation conditions as those of this experiment, and a Satiation (for approval and social contact) condition. The mean reinforcer effectiveness score was highest under Deprivation, intermediate under the Nondeprivation condition, and lowest under Satiation, unqualified by any interaction with sex of *S*s. And since the Satiation condition represents a logical extension of a dimension best characterized by a deprivation logic, the mean under this condition rank-orders consistently with the logic. These results give further support to the contention that social isolation represents essentially a deprivation of social reinforcers.

While these arguments seem to support the idea that social isolation is readily conceptualized as a deprivation of social reinforcers, considerable conceptual caution seems nevertheless to be in order. Probably the conclusion is best considered as tentatively supported pending data from more extensive parametric investigations of these variables.

Age and the effects of social deprivation. A final characteristic of deprivation as defined here is its greater effectiveness upon the older *S*s, evidenced by the positive relationship of Age with the Deprivation score. It is the older children in the sample for whom the social reinforcers of approval and attention are most significant, according to the reliable positive correlations with Age of the Approval and Attention ratings. Clearly, even in the narrow age ranged studied, it is the older children who may be expected to have the longer and better integrated histories of conditioning of these reinforcers. And it is not unreasonable to assume that children in this age range will have as yet only the beginnings of a history of conditioning of achievement as a reinforcer, a conclusion in accord with the zero-order relationship of the Achievement rating to the Deprivation score.

Sex and the effects of social deprivation. The behavior pattern of the boys with the female *E* under the two levels of social deprivation clearly is the primary determinant of the reliable second-order interaction of social deprivation with Sex of *S* and Sex of *E*. This interaction effect might be referred to a number of the rationales summarized and examined in (5) (the most prominent probably being Freud's Oedipus theory [4]). However, cautious interpretation is in order. The intent of the experiment was of course to exercise some control over the sex of the reinforcer-dispensing adult relative to the sex of the child; yet it must be observed that only one *E* was used to represent each sex. These two persons surely differed with respect to a great number of criteria, many of them of considerable theoretical pertinence. Thus, the reliable interaction effects in which the identity of the *E*s figure might be referred as reasonably to some of these possible differences as to sex. Only a wide replication of these interactions with suitably chosen *E*s can resolve this ambiguity. Still, the very similar cross-sex effect found in the Availability study cited (5) suggests that it is the sex patterns of *S*s and *E*s that are at issue.

SUMMARY

On the assumption that approval is representative of the reinforcers controlling the purely social initiations made by young children to adults, the verbal approval of an adult (*E*) was made contingent upon one response in a two-response game for 32 nursery school children (*S*s). The change in relative response frequency from a baseline level following introduction of adult approval, reflected the degree to which approval was reinforcing for an *S*. Adult approval was found to have definite reinforcing value for the *S*s ($p < .025$). With control exercised over sex of child, sex of adult, and sequence of conditions, the mean effectiveness of adult approval as a reinforcer for the *S*s was enhanced reliably, by a preceding period of 20 minutes of social isolation (conceived to be social deprivation), relative to its effectiveness for the same *S*s when not isolated ($p < .025$); and this result was qualified by the fact that the reinforcer dispensed by each *E* appeared more effective after social deprivation only with *S*s of the opposite sex, with this difference reliable alone for boys tested by the female (rather than the male) *E*. The increase in effectiveness of adult approval as a reinforcer produced in the children by social isolation was shown to be a positive function of the degree to which they characteristically sought adult approval in other settings ($p < .01$), and of age ($p < .001$), in the 46- to 63-month range studied.

Because social isolation excluded access to persons who might provide social reinforcers, its effect was referred to a concept of social deprivation, defined as the unavailability of social reinforcers. Thus it appears that the effectiveness of a representative social rein-

forcer may be enhanced by an operation of deprivation, as is the case for the reinforcers of primary appetitive drives like hunger and thirst. The effectiveness of this reinforcer appears enhanced particularly for *S*s who typically seek it.

REFERENCES

1. BOWLBY, J. Maternal care and mental health. *Bull. World Hlth Org. Monogr.*, 1951, **3**, 355–533.
2. BUTLER, R. A. The effect of deprivation of visual incentives on visual exploration motivation in monkeys. *J. comp. physiol. Psychol.*, 1957, **50**, 177–179.
3. CHASE, LUCILLE. Motivation of young children: An experimental study of the influence of certain types of external incentives upon the performance of a task. *Univ. Iowa Stud. Child Welf.*, 1932, **5**, No. 3.
4. FREUD, S. The passing of the Oedipus complex (1924). In *Collected papers*, Vol. 2. London: Hogarth, 1949. Pp. 269–276.
5. GEWIRTZ, J. L. Three determinants of attention-seeking in young children. *Monogr. Soc. Res. Child Develpm.*, 1954, **19**, No. 2 (Serial No. 59).
6. GEWIRTZ, J. L. A program of research on the dimensions and antecedents of emotional dependence. *Child Develpm.*, 1956, **27**, 205–221.
7. GOLDFARB, W. Effects of psychological deprivation in infancy and subsequent stimulation. *Amer. J. Psychiat.*, 1945, **102**, 18–33.
8. GREENSPOON, J. The reinforcing effect of two spoken words on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
9. GUILFORD, J. P. *Fundamental statistics in psychology and education.* (2nd ed.) New York: McGraw-Hill, 1950.
10. HARTUP, W. W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of preschool children. Unpublished doctoral dissertation, Harvard Univer., 1955.
11. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
12. HURLOCK, ELIZABETH B. The value of praise and reproof as incentives for children. *Arch. Psychol., N. Y.*, 1924, **11**, No. 71, 5–78.
13. KELLER, F. S., & SCHOENFELD, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
14. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
15. LUBIN, A. Some formulae for use with suppressor variables. *Educ. psychol. Measmt*, 1957, **17**, 286–296.
16. ORTMEYER, D. H. Drive generalization from hunger to fear. *Dissertation Abstr.*, 1955, **15**, 1257–1258. (Abstract)
17. RIBBLE, MARGARET. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 621–651.
18. SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychologist*, 1951, **6**, 476–483.
19. SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.
20. SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. In Anna Freud *et al.* (Eds.), *The psychoanalytic study of the child.* Vol. 1. New York: Int. Universities Press, 1945. Pp. 53–74.
21. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.
22. WOLF, THETA S. The effect of praise and competition on the persisting behavior of kindergarten children. *Univer. of Minnesota Inst. Child Welf. Monogr. Ser.*, 1938.

Received March 27, 1957.

PREFERENCE FOR DELAYED REINFORCEMENT: AN EXPERIMENTAL STUDY OF A CULTURAL OBSERVATION[1]

WALTER MISCHEL
University of Colorado

This study is, in part, an attempt to validate, experimentally, an observation about cultural differences formulated during the course of a "culture and personality" field research in Trinidad, British West Indies. It may be of interest, therefore, not only for its specific findings but also as an example of experimentation in the cultural field.

On the basis of anthropological observation, a major personality difference between the Negro and East Indian ethnic groups of the island of Trinidad was suggested. This difference, as expressed by numerous informants, is that the Negroes are impulsive, indulge themselves, settle for next to nothing if they can get it right away, do not work or wait for bigger things in the future but, instead, prefer smaller gains immediately. In contrast to this, the Indian is said to deprive himself and to be willing and able to postpone immediate gain and pleasure for the sake of obtaining greater rewards and returns in the future. In effect, when given a choice, the Negro is said to be characterized by preference for relatively smaller, immediate reinforcements, whereas the Indian is said to prefer larger, delayed reinforcements.

Judgments centering about such a distinction of differential preference or tolerance for delayed reinforcement are, of course, also quite prevalent within our own culture. The widely observed preferences of children for immediate rewards has been commented on (2, 3, 5) and is reflected in some theoretical formulations, e.g., the "pleasure principle" and the "reality principle" (4, 6). In clinical applications, the inability to postpone immediate gratification for the sake of delayed rewards is often considered an important factor in immaturity, maladjustment, and "psychopathy." Mowrer and Ullman (10), for example, consider the delay of reinforcements to be a critical factor in the development of neurotic and criminal behavior.

Although the research on delay of reinforcement in human conditioning has been vast (Wolfle's studies [14, 15] are representative), there has been comparatively little experimentation with human subjects on variables other than the effect of the time interval itself on learning rate or object preference (e.g., 1, 7, 8, 11, 13) and outside the conditioning paradigm. A notable exception is Mahrer's (9) attempt to apply Rotter's expectancy construct (12). Mahrer was able to make successful predictions about children's choices for immediate versus delayed reinforcements as a function of differential expectancies for obtaining positive reinforcement following delay. That is, the expectancy that reinforcement would follow from the social agent even after time delay was shown to be an important variable. In common sense terms, this kind of behavior may be thought of as "trust" or the belief that the agent promising delayed reinforcement will actually supply it. The emphasis in Mahrer's study, and in the present research, is on the expectancy for reinforcement to follow as a function of making a particular choice with a particular social agent in a given situation, rather than on the time interval itself.

One may assume that expectancies for reinforcement to follow from certain social agents, in spite of time delay, are learned in a manner fundamentally similar to the learning of any social behaviors, and are governed by similar principles of generalization. It seems probable, then, that a child's expectancies that delayed reinforcement will issue from new adult social agents are related to his past experiences in which promised reinforcement followed delay from such major adult sources of learning as the parental figures within the home. In a situation in which a child is confronted with a new male social agent, we would anticipate that his expectancies for promised reinforcement to follow from that social agent, in spite of time delay, are related to past experiences of this kind with the father or father figure in the home. Similarly, we would anticipate that the child's expectancies that promised reinforcement will follow, in spite of delay, from female social agents are related to past experiences of this kind with the mother or mother figure in the home. (Since in the present experiment the social agent (E) is male, our discussion and prediction center about expectancies for delayed reinforcement from male social agents.) We would further expect children who have had less of such experience with a male social agent in the home to show less "trust," or lower expectancy for reinforcement to follow in spite of time delay when promised by a new male social agent, than children who have had more of such experience. Presumably, a child whose father is absent in the home has had more limited experiences in which promised reinforcement follows in spite of time delay from a male social agent. We anticipate that such a child would have lower expectancy for delayed reinforcement from a male social agent than a child whose father is present within the home. Specifically, we would expect the "fatherless" child to show less preference for a larger but delayed reinforcement and more preference for a smaller but immediate reinforcement than the child whose father is present within the home.

---

[1] This study was conducted while the author collaborated with Frances Mischel in an interdisciplinary project which was sponsored by the Wenner-Grenn Foundation for Anthropological Research. A report based on portions of this paper was read at the convention of the Rocky Mountain Psychological Association, April, 1957.

The above prediction seemed, on the basis of tentative observation, to apply differentially to the two cultural groups we are considering; namely, examination of the family structures of the Negro and Indian groups within the particular culture and locale suggested that the Negro family seemed more frequently characterized by the absence of a temporally stable adult male figure in the home than was the Indian family. The presence or absence of a father within the family constellation was selected as a key variable influencing choice behavior with respect to delayed or immediate reinforcement and the aim of the experiment was: (*a*) to test the hypothesis of differential preference for larger, delayed as opposed to smaller, immediate reinforcement of the two ethnic groups, it being predicted that the Indian group would show greater preference for larger, delayed reinforcement than the Negro group; and (*b*) to test the hypothesis of differential preference for larger, delayed as opposed to smaller, immediate reinforcement, issuing from a male social agent, of children whose fathers are present or absent within the home, it being predicted that children having a father or father figure present within the home would prefer the larger, delayed reinforcement more than children lacking such a figure. Further, since age and socioeconomic level seemed potentially relevant variables for this kind of choice behavior, this information was included in the data collected, with the intent of relating it to the choice.

METHOD

Subjects

The subjects (*S*s) included 53 male and female children between the ages of 7 and 9 in the elementary section of a rural Trinidad school. The school was located in the southern part of the island and seemed representative of rural schools within the colony. Thirty-five of the children were Negro and eighteen East Indian. The sex ratio was 29 males to 24 females. The children of the two ethnic groups attended the same large section or class and it was thus possible to administer the experimental test to the members of both ethnic groups at the same time.

The Reinforcements

Pre-experimental sessions were required to select two reinforcements of use for the experimental situation. Towards this end, 15 Negro and Indian boys and girls, representing the above ages, and selected from another rural Trinidad school, were seen in individual sessions and their preferences for various specific reinforcements were elicited. As a result, two reinforcements (both candy, but varying markedly in size, price, and packaging, i.e., a one-cent and a ten-cent candy) were selected. These met the desired requirements inasmuch as the larger reinforcement was uniformly preferred in a straight choice situation ("which *one* of these two would you like to take"), but when the choice was "you can have this one (the smaller) today *or* this one (the larger) in one week" approximately 50 per cent of the group chose the former and approximately 50 per cent the latter.

Task and Measures

The *S*s' task was to fill out a simple questionnaire. All the *S*s were sufficiently literate for this. The questionnaire items of experimental interest were ethnic group, age, socioeconomic status, and presence or absence of the father within the household. Ethnic group was measured by *S*'s checking either "Creole" or "Indian," and age by a "fill in" response. The measure of socioeconomic status was quite gross, and *S* merely checked "galvanized" (tin roof, comparatively expensive and generally a sign of relatively high socioeconomic status in the rural areas) or "carrat" (thatched leaf roof, inexpensive, and generally a sign of low socioeconomic status), in response to an item concerning the kind of house he lives in. This measure provided a crude dichotomization into relatively "high" and relatively "low" socioeconomic status. "Presence" or "absence" of the father in the home was measured by the response "yes" or "no" to the questions "does your father live at home with you," and "do your mother *and* father live at home with you." One *S* reported that father lived at home, mother lived at home, but that mother and father did not live at home and was eliminated from the analysis for this inconsistency. All but three *S*s reported "yes" to the question "does your mother live at home with you." These three exceptions lived with their grandmothers and because of this deviation were omitted from the statistical analyses.

An informal check on the accuracy of the children's responses was attempted with the headmaster who was familiar with the home situations of most of the *S*s. This check was undertaken on a sample of twenty of the protocols and indicated that, to the headmaster's best knowledge, the responses corresponded with his own information about the children.

Procedure

The experimenter (*E*) was introduced as an American interested in gathering information on children in the local schools of the island. To help with this the *S*s were asked to fill out the questionnaire. When these were completed, *E* expressed his wish to thank the group for their cooperation. He displayed the two kinds of reinforcements and said: "I would like to give each of you a piece of candy but I don't have enough of these (indicating the larger, more preferred reinforcement) with me today. So you can either get this one (indicating the smaller, less preferred reinforcement) right now, today, or, if you want to, you can wait for this one (indicating) which I will bring back next Wednesday (one week delay interval)." To insure clarity, these instructions were repeated in rephrased form and both reinforcements were carefully displayed. The fact that getting the (smaller) candy today precluded getting the (larger) one next week, and vice versa, was stressed. *S*s were asked to indicate their choice by writing "today" (T) or "next week" (W) on their questionnaires. The response made here was the measure of choice of a larger (or more preferred) delayed reinforcement or a smaller (less preferred) immediate reinforcement. *S*s were seated sufficiently far apart from each other to insure reasonably that their choices were made independently in this group setting.

RESULTS AND DISCUSSION

The data from the present experiment were analyzed in terms of differential preference for smaller, immediate reinforcement versus larger, delayed reinforcement in relationship to ethnic group, presence or absence of the father within the home, age, and socioeconomic status. All statistical analyses were made with the chi-square test. The data used for relating the experimental choice with ethnic group and presence or absence of the father within the home, are given in Table 1.

TABLE 1

RELATIONSHIPS BETWEEN ETHNIC GROUP, PRESENCE OF THE FATHER, AND EXPERIMENTAL CHOICE

Group	*N* Choosing Immediate Reinforcement	*N* Choosing Delayed Reinforcement	χ^2	p
Negro	22 (18.49)[a]	13 (16.51)	$\begin{cases} 4.17 \\ 3.06^b \end{cases}$	<.05 <.10
Indian	6 (9.51)	12 (8.49)		
Father present	17 (22.19)	25 (19.81)		
Father absent	11 (5.81)	0 (5.19)	10.13^b	<.01
Negroes with father present	12 (15.71)	13 (9.29)		
Negroes with father absent	10 (6.29)	0 (3.71)	6.19^b	<.02

[a] Figures in parentheses are expected frequencies.

[b] Chi square corrected with Yates correction, for continuity.

Comparison of the Negro and Indian groups on the experimental choice (Table 1) yielded a chi square of 4.17 (p between .05 and .02), a significantly larger proportion of the Negro *S*s choosing immediate reinforcement. It should be noted, however, that if the Yates correction is applied (the expected frequencies within two of the cells are between eight and ten), the chi square is reduced to 3.06 having a p value of less than .10. Nevertheless, this difference between the ethnic groups on the experimental choice still seems sufficient to require caution in combining them on other analyses unless it can be shown that members of the two groups are in comparable proportions on the other relevant variables.

The comparison between ethnic groups with respect to presence or absence of the father does not indicate a difference at an acceptable probability level (chi square is 2.56 with p $< .20$ and $> .10$). Some ancillary evidence supporting the trend of differential fatherlessness within the two ethnic groups can be drawn from another source. Data, including the measure of presence or absence of the father within the home, were collected within the same school for older children, of both ethnic groups, who were not included in the present study. If the data from these children ($N = 24$) on presence or absence of the father for the two ethnic groups are combined with data from the *S*s of the present study, a chi square of 5.57 is obtained which is significant beyond the .02 level, and in the direction expected from cultural observation.

In view of these findings, the relationship between presence or absence of the father and the experimental choice was tested for each of the ethnic groups separately as well as for the total combined sample. As there is only one "fatherless" *S* in the Indian group, such a test is not meaningful for that group. For the Negro group alone, and for the total sample, the comparison between presence or absence of the father and the experimental choice (Table 1) yielded chi squares of 10.13 and 6.19 respectively, which are significant beyond the .01 and .02 levels, and indicate, as predicted, a greater proportion of choices for immediate reinforcement on the part of "fatherless" *S*s.

Before this relationship between presence of the father and preference for delayed, larger reinforcement can be generalized, it will of course be necessary to test other cultural groups in a similar experimental situation. The present research emphasized the relationship only between the "physical" presence of *S*'s father in the home and preference for delayed reinforcement issuing from another male social agent. In view of the obtained significant finding, the relationship between the "psychological" presence of the father (e.g., degree of identification) and this kind of behavior merits further investigation. Such research would seem especially relevant for a further understanding of impulsive behavior and "psychopathy."

The presence or absence of the father may be significant not as, or just as, a causal factor in itself in relation to preference for delayed versus immediate reinforcement, but as an index of the extent to which the parents or the family participate in a delayed reward culture. That is, the presence of the father in the home (in contrast, for example, to a ready abandoning of the home) may, in turn, be related to the degree of family participation in a cultural pattern in which delayed, larger reinforcement is preferred or prized more than immediate, smaller reinforcement. Such participation may have many other impacts upon the child apart from, or in addition to, the paternal presence.

Neither the proportion of Negro and Indian *S*s, nor the proportion of *S*s with father present and absent, was significantly different for the three age groups or for their socioeconomic status groups (both *p* levels greater than .20). Therefore, tests of the relationship between age and the experimental choice, and between socioeconomic status and the experimental choice, can be made with all *S*s combined. Table 2 shows the number of *S*s in each age group choosing immediate versus delayed reinforcement. The chi-square test applied to this table yields a corrected chi-square value of 7.10 which is significant beyond the .05 level and indicates an increasing proportion of *S*s choosing delayed reinforcement at the older age levels.

TABLE 2

RELATIONSHIP BETWEEN AGE AND EXPERIMENTAL CHOICE

Experimental Choice	Age			χ^2	p
	7	8	9		
N Choosing immediate reinforcement	13 (8.45)[a]	13 (14.27)	2 (5.28)		
N Choosing delayed reinforcement	3 (7.55)	14 (12.73)	8 (4.72)	7.10[b]	<.05

[a] Figures in parentheses are expected frequencies.

[b] Chi square corrected with Yates correction, for continuity.

This relationship between *S*s' age and preference for delayed reinforcement is in accord with nonexperimental discussions of such a relationship to the effect that with increasing maturity comes the increasing ability to delay gratification. The finding is consistent with an interpretation of preference for delayed reinforcement as a learned behavior which is, in part, a function of the expectancy that the promised reinforcement will issue from the social agent in spite of time delay. With increasing age, the potentiality for developing a strong expectancy of this kind increases *if* the individual continues to gain reinforcing experiences within this area, thus building up the relevant expectancies, but not as a function of growing older or biological maturation per se.

Comparison of the "high" versus "low" socioeconomic groups on the experimental choice did not yield a significant difference (*p* value greater than .20). This lack of established relationship between socioeconomic level and the experimental choice cannot be considered definitive in view of the extreme crudeness of the measure of socioeconomic level employed.

SUMMARY

This experiment tested an observation, made by anthropological field techniques, regarding personality differences between the East Indian and Negro populations of Trinidad, B.W.I. The primary results were significant differences between preference for immediate, smaller versus delayed, larger reinforcement and, first, the presence or absence of the father within the home, and second, age. The over-all findings appear to have implications not only for the two specific ethnic groups studied here but also for further research on relationships between personality variables of a less gross kind and this type of choice behavior within our own culture. If there are relationships between "maturity" or "adjustment" and tolerance, or preference, for larger, delayed as opposed to smaller, immediate reinforcement, then the isolation of major variables related to such preferences seems particularly important.

REFERENCES

1. BRENNER, B. Effect of immediate and delayed praise and blame upon learning and recall. *Teach. Coll. Contrib. Educ.*, 1934, No. 620.
2. COLE, LUELLA, & MORGAN, J. J. B. *Psychology of childhood and adolescence.* New York: Rinehart, 1947.
3. FAEGRE, MARION L., & ANDERSON, J. E. *Child care and training.* Univer. of Minnesota Press, 1937.
4. FREUD, S. *Beyond the pleasure principle.* New York: Boni & Livewright, 1922.
5. HARRIS, D. B. How children learn interests, motives, and attitudes. In N. B. Henry (Ed.), *49th Yearbook, NSSE, Part I, Learning and instruction*, 129–155.
6. HENDRICK, I. *Facts and theories of psychoanalysis.* New York: Knopf, 1948.
7. IRWIN, F. W., ARMITT, F. M., & SIMSON, C. Studies in object preferences: I. The effect of temporal proximity. *J. exper. Psychol.*, 1943, **33,** 64–72.
8. IRWIN, F. W., ORCHINIK, C. W., & WEISS, JOHANNA. Studies in object preferences: The effect of temporal proximity upon adults' preferences. *Amer. J. Psychol.*, 1946, **59,** 458–462.
9. MAHRER, A. R. The role of expectancy in delayed reinforcement. *J. exp. Psychol.*, 1956, **52,** 101–105.
10. MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, **52,** 61–90.
11. NAGGE, J. W., *Psychology of the child.* New York: Ronald, 1942.
12. ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.
13. SALTZMAN, I. J., Delay of reward and human verbal learning. *J. exp. Psychol.* 1951, **41,** 437–439.
14. WOLFLE, H M. Time factors in conditioning finger withdrawal. *J. genet. Psychol.*, 1930, **4,** 372–378.
15. WOLFLE, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. genet. Psychol.*, 1932, **7,** 80–103.

Received March 25, 1957.

EMPATHY OR BEHAVIORAL TRANSPARENCY?[1]

URIEL G. FOA

The Israel Institute of Applied Social Research and Tel Aviv University

The term social empathy has been generally used to indicate the activity of guessing another person's perception with regard to the behavior or values of a given individual. For example, A guesses B's perception of C's behavior or values. It is generally understood that A and B are two different individuals. Individual C, however, may be either a third person distinct from A and B, or identical with either A or B, as in a situation involving two people only. It seems quite possible that A might succeed in guessing B's perception of the behavior or values of B, and fail to guess B's perception of the behavior or values of A.

The following four aspects or facets of empathy may therefore be worth considering: (*a*) the guesser, or the individual who guesses the perception of another person; (*b*) the guessee, or the person whose perception is guessed; (*c*) the actor, or the individual whose behavior or values in a particular role are guessed about; and (*d*) the level, which indicates whether the guessing regards the actual behavior of the actor, or the corresponding values.

Most of the previous studies of empathy have been concerned with the guesser and/or the guessee. The actor has been tacitly assumed to be identical with the guessee and, in general, he does not appear explicitly in the design of the experiments reported in the literature. For some investigators, empathy is a personality trait of the guesser that is possessed in different degrees by different subjects. Tests of empathy (2, 5) and studies of the relationship between empathy and other personality traits (8) stem from this conception.

Other students of empathy have held that this variable depends on guesser and guessee alike: if A can predict one aspect of the response of B, he can also predict other aspects of B's response, although he might fail in the prediction of the responses of another person (7).

A third approach is suggested by Tagiuri and others (9). Essentially, it consists in looking at empathy from the vantage point of the guessee rather than from the point of view of the guesser. People, says Tagiuri, are more or less easy to understand; they possess a different degree of transparency. The feelings of transparent people are easier to guess. Empathy scores depend on the guessee.

It is now suggested that success or failure in guessing correctly the perception of somebody else may depend not only on the guesser and/or the guessee but also on the actor. If the actor behaves in an unambiguous manner, it may prove easy to guess the perception that somebody gets of this behavior. Vice versa, if the behavior of the actor is ambiguous, people may get blurred perceptions of it and the guessing of these perceptions may be difficult.

The hypotheses to be tested can be formulated in the following way:

Hypothesis I. Guesses made by the same guesser with regard to different guessees correlate among themselves. For example: if Bill is a good empathizer he will guess correctly the perceptions of other people, no matter who the guessees and the actors are. This is the personality trait approach.

Hypothesis II. Guesses made by different guessers with regard to the same guessee correlate among themselves. This is the transparency approach.

Hypothesis III. Different guesses made by the same guesser with regard to the perceptions of the same guessee correlate among themselves.

Hypothesis IV. Guesses made by different guessers with regard to different guessees correlate among themselves if they all deal with the behavior of the same type of actor. This is the approach introduced in the present paper.

Hypothesis III combines Hypotheses I and II. It is therefore evident that if Hypothesis III is rejected then Hypotheses I and II are also rejected. The acceptance of Hypothesis III, however, does not necessarily prove Hypotheses I and II. The experiment described in this paper permits the testing of Hypotheses III and IV. These two hypotheses are not mutually exclusive; that is the acceptance of one of them does not lead to the necessary rejection of the other.

---

[1] This paper is one in a series of studies on interpersonal relations in industry which have been initiated with partial support from the Ford Foundation, and was read at the 1957 Annual Meeting of the Israel Psychological Association.

METHOD

The Test of Foreman-Worker Relationship (4) was administered to 490 Israeli factory workers and their 51 foremen. This test consists of two sets of pictures of eight situations each. One set is concerned with the behavior of the foreman and the other with the behavior of the worker. Both sets of situations are presented to the foreman and the worker. For each situation they are instructed to choose, out of four given types of behavior, the one which (*a*) ought to occur in an ideal situation, and (*b*) occurs most often in practice in their department. Further, each worker is requested to guess the alternatives chosen by his own foreman, both at the "ought to be" and the actual level. Similarly, each foreman is requested to guess the most frequent (or modal) choice of his workers at both levels.

Let us consider a dyad composed of a worker and his foreman. Each of the two subjects (*S*s) of the dyad gives four sets of responses of eight responses each, thus:

1. Responses regarding the *ideal* behavior of the *foreman*.
2. Responses regarding the *actual* behavior of the *foreman*.
3. Responses regarding the *ideal* behavior of the *worker*.
4. Responses regarding the *actual* behavior of the *worker*.

Further, each *S* guesses the responses of the other.

The *empathy* score for each set of eight responses is computed by counting the number of times the guess of one *S* is the same as the corresponding response of the other. Since there are eight responses in each set, it is evident that the maximal empathy score is eight, and obtained when the *S* has succeeded in guessing correctly all the eight responses of the set and the minimal score zero. Each *S* guesses four sets of responses and thus receives four empathy scores. In a dyad composed of a worker and his foreman there are eight empathy scores, of which four pertain to the worker and four to the foreman.

A measure of *real similarity* in response for each set of eight pictures was computed by counting the number of times the response of one *S* is the same as the response of the other *S*. A high score indicates that the two *S*s tend to have the same perception of a given type of behavior.

In the population of dyads, the product-moment coefficient of correlation was then computed for each possible pair of empathy scores. Since for each dyad there are eight scores, 28 pairs and their respective coefficients, are possible. Each coefficient indicates the correlation between two empathy scores which differ from one another in at least one of the following elements: (*a*) the *S* with whose behavior the empathy is concerned: either the worker or the foreman; (*b*) the *S* making the guesses and/or the *S* guessed about (in a dyad, when the worker is the guesser, the guessee is the foreman and vice versa); and (*c*) the level of behavior with which the empathy is concerned: either actual or ideal.

When a particular coefficient differs significantly from zero, this suggests that at least one of the elements that are identical in the two correlated scores may be important in the empathic process. On the other hand, when a coefficient approaches zero, this suggests that the elements identical in the two scores are not relevant to empathy in the situation being studied. This kind of reasoning is somewhat similar to factor analysis. In factor analysis, one infers the existence of a common factor when two variables are correlated. Here, each variable is described a priori in terms of certain factors, and the eventual correlation between two variables is attributed to the factors they have in common.

RESULTS

The 28 product-moment correlation coefficients for each possible pair of empathy scores, are reported in Table 1. At the .05 level, coefficients of the correlation between two empathy scores for workers are significant when above .09; coefficients of the correlation between two scores for foremen are significant when above .27. For coefficients between a score of the worker and a score of the foreman the critical value is, of course, somewhere between these two values. An approximation to this value is given by the formula: $F = r^2 \Big/ \dfrac{1 - r^2}{H\ (N-2)}$, where F is the variance ratio for the .05 probability level; N the number of foremen; and H the harmonic mean of the number of workers per foreman. In the present case, the formula shows that coefficients of correlation above .15 are significant.

In Table 1, there are twelve coefficients between scores referring to the behavior of the same type of actor (upper left and lower right quadrants of the table); eight of them are significant. On the other hand, no significant coefficient is found among the 16 correlations between scores referring to different types of actor (lower left quadrant of the table). For instance: the empathy score of the worker for the norms of the foreman regarding the behavior of the worker correlates at .32 with the empathy score of the foreman regarding the worker's perception of the actual behavior of the worker. Here the actor is the same, but guesser, guessee, and level are different, and yet the correlation is fairly high. On the other hand, the empathy of the workers for

TABLE 1
EIGHT EMPATHIES: INTERCORRELATIONS AND AVERAGE SCORES

Form of Empathy	Correlation Coefficient								
		1	2	3	4	5	6	7	8
Behavior of foreman:									
Worker's guesses of foreman's responses									
Actual level	1	—							
Ideal level	2	.34	—						
Foreman's guesses of worker's responses									
Actual level	3	.26	.19	—					
Ideal level	4	.22	.13	.15	—				
Behavior of worker:									
Worker's guesses of foreman's responses									
Actual level	5	.04	.06	.06	.15	—			
Ideal level	6	.04	.06	−.04	.05	.26	—		
Foreman's guesses of worker's responses									
Actual level	7	.13	.02	.03	−.01	.29	.32	—	
Ideal level	8	.15	−.09	.03	−.03	.06	.27	.34	—
Correlation with real similarity	—	.42	.25	.58	.58	.56	.37	.57	.44
Average empathy score	—	3.44	3.85	3.26	4.24	3.32	4.04	4.08	4.46

the foreman's perception of the foreman's actual behavior correlates at .04 with the empathy of the worker for the foreman's perception of the worker's actual behavior. Here the actor is different, but guesser, guessee, and level are the same in both scores, and the correlation is practically zero. The average coefficient between scores with the same type of actor is .24: the correlation is significant. The average coefficient between scores with a different type of actor is .05, which is below the critical value. These results seem to support the acceptance of Hypothesis IV.

The average correlation between scores with the same type of guesser and guessee is .13. The average correlation between scores with a different type of guesser and guessee is .14. The two coefficients are practically alike and below the critical value, suggesting that Hypothesis III should be rejected. Rejection of Hypothesis III leads, in turn, to the rejection of Hypotheses I and II.

The next to the last row of Table 1 gives the coefficients of correlation between each type of empathy score and the corresponding score of real similarity. These coefficients indicate the extent to which the accuracy of the *S*, in guessing the perception of the other, is related to the fact that both have the same perception. Let us think, for example, of a number of *S*s who always attribute to the other their own perception. Then their guesses would be right when the two perceptions are alike, and wrong when the two perceptions are different. In this case the investigator would find perfect positive correlation between empathy and real similarity scores. In our case, the coefficients of correlation between real similarity score and the corresponding empathy score show that accuracy of guessing is, to some extent, a function of real similarity. The two variables, however, are far from being identical.

In the last row of Table 1 one finds the average score for each type of empathy. The significance of the differences between any two of these averages was checked by the *t* test for correlated pairs and the following results were obtained:

1. The average score at the actual level is always significantly lower than the corresponding score at the ideal level. The significance level is .001 in three cases and .05 in one case.

2. The average scores referring to the empathy of the foreman are better than the corresponding scores referring to the empathy of the worker, with one exception: the score of the foreman regarding the foreman's actual behavior. All these differences are significant; three of them at the .001 level and one at the .05 level.

3. The average scores referring to the behavior of the worker are better than the corresponding scores referring to the behavior of the foreman, again with an exception: the score of the worker regarding the worker's actual behavior. Differences between scores regarding ideal behavior are significant. Those regarding actual behavior are not.

4. The two worst scores are the one of the foreman regarding the foreman's actual behavior and the one of the worker regarding the worker's actual behavior. The difference between each one of these average scores and all the other averages is significant respectively in five and four cases out of six.

DISCUSSION

The findings suggest the following conclusions:

1. The accuracy of A in predicting a certain aspect of the social perception of B is not necessarily related to A's accuracy in predicting another aspect of the perception of B.

2. Empathy scores referring to the perception of the behavior of the same actor are likely to be moderately correlated among themselves.

3. Empathy is partly determined by similarity of perception between guesser and guessee.

It seems possible that the relationship between empathy scores referring to the same type of behavior may be due to the degree of similarity of perception of such behavior. Indeed, any two empathy scores referring to the same level and the same actor are both related to the same similarity score. The real similarity score, for a given level and a given actor, is common to the worker and to his foreman. This suggests that the relationship between empathy scores referring to the same actor might be due to the similarity between the perceptions of the guesser and of the guessee. Indeed, when the common real similarity score is partialled out, the correlation coefficients between empathy scores tend to vanish.

It seems, therefore, that unambiguous behavior is more likely to be perceived in the same manner by different people than ambiguous behavior. If the behavior to be guessed varies considerably in ambiguity, as in the present case, it may influence the relationship between empathy scores. As a consequence, it may be suggested that the type of behavior to be guessed is one of the variables that should be taken into account in the design of studies of social perception and empathy. Failure to control this variable can introduce a source of unexplained variance. Keeping it constant, on the other hand, may help the investigator in determining the portion of the variance of empathy scores which is related to the guesser and/or the guessee.

The comparison of the average empathy scores seems to confirm a previous finding by Nagle (6) that ego-involvement impairs empathy: both foremen and workers are indeed at their worst when guessing the other's perception of the behavior of their respective roles.

The finding that the foreman is a better empathizer than the worker is in agreement with the results reported by Bell and Hall (1) who found that leadership and empathy are related.

On the other hand, our data suggest that values regarding the behavior of the worker are more transparent than the values regarding the behavior of the foreman. Also, they show that it is easier to guess a person's values than it is to guess his perception of actual behavior.

SUMMARY

This is a study of the intercorrelations among eight different kinds of empathy between 490 industrial workers and their respective foremen.

The findings show that measures of empathy referring to the behavior of the same individual are correlated. On the other hand, those referring to the behavior of different individuals are uncorrelated, even when the guesses are made by the same subject and/or with reference to the responses of another given subject.

The present data are thus in disagreement with theories that empathy depends on the personality of the guesser or on the personality of the guessee. They suggest that when people behave in a transparent manner it is easier to guess the perception of this behavior; that is, empathy is related to transparency of behavior. The type of behavior is therefore a variable that should not be disregarded in the study of empathy.

The average scores on the eight kinds of empathy are also discussed. Guessing of the other's perception of the behavior of one's own role is of particular difficulty for both the foreman and the worker. In general, the empathy scores of the foreman are better than the corresponding scores of the worker. Guessing the other's values is easier than guessing his perception of actual behavior.

REFERENCES

1. BELL, G. B., & HALL, H. E. The relationship between leadership and empathy. *J. abnorm. soc. Psychol.*, 1954, **49**, 156–157.
2. BENDER, I. E., & HASTDORF, A. M. On measuring generalized empathic ability (social sensitivity). *J. abnorm. soc. Psychol.*, 1953, **48**, 503–506.
3. FOA, U. G. The Foreman-worker interaction: A research design. *Sociometry*, 1955, **18**, 226–244.

4. FOA, U. G. A test of the foreman-worker relationship. *Personnel Psychol.*, 1956, **9**, 469–486.
5. JARRARD, L. E. Empathy: The concept and industrial applications. *Personnel Psychol.*, 1956, **9**, 157–167.
6. NAGLE, B. F. Productivity, employee attitude and supervisor sensitivity, *Personnel Psychol.*, 1954, **7**, 219–233.
7. PATTON, W. M., JR. Studies in industrial empathy: III. A study of supervisory empathy in the textile industry. *J. appl. Psychol.*, 1954, **35**, 285–288.
8. TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **51**, 1–23.
9. TAGIURI, R., KOGAN, N., & BRUNER, J. S. The transparency of interpersonal choice. (Stencilled), Laboratory of Social Relations, Harvard Univer., n.d.

Received April 2, 1957.

RIGIDITY, INSECURITY, AND STRESS[1]

LEONARD H. AINSWORTH[2]
University College, London

THE controversy as to whether rigidity is a situational phenomenon or a generalized personality characteristic is the context from which the present investigation has arisen. Luchins (11, 12, 13, 14), whose water-jar Einstellung test has been used frequently as a measure of rigidity, has steadfastly emphasized situational factors as determinants of rigid performance in problem-solving situations. On the other hand, Rokeach (15, 16) and many others have sought to establish that rigid behavior in the problem-solving situation is attributable to a generalized personality characteristic that predisposes the individual to rigid attitudes and behavior in a broad variety of situations. It has been proposed that rigidity emerges as a defense against anxiety or as a result of defenses against anxiety. Some authors (e.g. 7, 8) imply that chronically anxious people tend to manifest generalized rigidity, while others (e.g. 18) view as primarily situational both the anxiety and the rigidity which serves to reduce it. Levitt (10), in reviewing the literature of the water-jar Einstellung test, finds so little consistency in the findings of many investigations that he concluded that the test is not a valid measure of any generalized trait of rigidity. Commenting on the influence of situational stress upon performance on the water-jar test, he says that "the over-all picture is beclouded."

The general hypothesis underlying this investigation is that rigidity is attributable to insecurity, and, more specifically, that problem-solving rigidity is a function of insecurity experienced in the problem-solving situation. Although on the surface this hypothesis appears to imply that rigidity is situational rather than personality-related, in fact insecurity is offered as a bridge-concept between the two extreme and contradictory views that have heretofore predominated.

The view of insecurity, stress, and rigidity that underlies the present investigation may be outlined as follows. Insecurity may be induced in an individual by the stressful aspects of a given situation, but the degree of insecurity he experiences is not directly proportional to the degree of stress as objectively estimated. First, the train of symbolic processes aroused in a previous insecurity-producing situation may prolong the experience of insecurity beyond the limits of the particular situation in which it was aroused, so that the individual enters a new situation already feeling insecure. Secondly, a situation that is nonstressful to many individuals may be stressful to others, either because they are relatively ill-equipped to deal with it, or because they perceive the situation as stressful through some similarity it bears to other situations in which insecurity was previously experienced.

Some individuals may be described as characteristically more insecure than others because they are insecure in a greater number and variety of situations. It seems reasonable that such an "insecure" person would be more likely than a relatively "secure" person to enter a given problem-solving situation in an insecure state, or to be made insecure by the demands of the task even under relatively nonstressful conditions. It is because insecurity may be viewed both as situational and as personality-related that it can serve a mediating function between the two apparently contradictory views of rigidity.

Rigidity is believed to arise as a defense against the insecurity experienced in a specific situation, but this does not imply that rigidity in this specific situation is related to a general ego-defensiveness, save through the intervening variable of insecurity. Thus it is not expected that a person with a "rigid personality structure," or one who makes massive use of defense mechanisms of various sorts, will be rigid in a problem-solving situation unless he experiences insecurity therein. To the extent that his general defenses prevent him from experiencing insecurity in the specific situation, he may perform flexibly.

---

[1] This report is an abridgement of a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of London. The author wishes to express his appreciation to his advisor, Dr. Roger W. Russell, for his helpful criticisms in the design of the experiment and for his advice throughout the original study.

[2] Now with the Medical Service, Supreme Bench, Baltimore, and the Johns Hopkins University.

In order to explore the validity of this general position, the following specific hypotheses were formulated:

1. The more insecurity that the individual experiences in his general life adjustment, the more likely he is to behave rigidly in a problem-solving situation. The assumption here is that the individual may carry insecurity over from other situations to the problem-solving situation.

2. The experimental introduction of stress in the problem-solving situation will increase the likelihood that the individual will behave rigidly therein, and the greater the stress the greater is the likelihood that he will behave rigidly. The assumption here is that the greater the stress the more insecure the individual will feel in the specific situation.

3. Defense mechanisms developed in general life adjustment are not directly related to rigid behavior in the problem-solving situation. To the extent that they are related to rigidity it is through insecurity operating as specified in the first hypothesis.

4. The introduction of stress tends to act differentially, according to the extent to which insecurity is experienced by the individual in his general life adjustment. This hypothesis implies that the two independent variables of insecurity and stress act together in affecting the dependent variable of rigidity.

METHOD

Criteria of Rigidity

The criteria of rigidity employed in this study are based on persistence in employing an Einstellung formula in the problem-solving situation presented by a water-jar test similar to that designed by Luchins (11). The general features of the test have been described in the literature and will not be repeated here. This particular version of the test consists of twelve problems. Problems 1 to 6 are Einstellung problems that can be solved only by employing the relatively complex Einstellung formula (B − A − 2C). Problems 7 to 9 are known as critical problems that may be solved either by the Einstellung formula or in a simple and direct way. Problems 10 to 12 are extinction problems that can be solved simply and directly but not by the Einstellung formula.

Luchins (12, 13, 14) insists that the proper criterion of problem-solving rigidity is the use of the Einstellung formula in extinction problems for which it provides an incorrect solution. The critical problems do not require a shift of method, since the Einstellung formula provides a correct solution, and therefore, he argues, the person who uses the Einstellung formula in the solution of the critical problems is not necessarily rigid. Other workers have defined rigidity as the use of the Einstellung formula in the solution of the critical problems, and have measured rigidity in terms of the number of critical problems that were solved by means of this formula. Rokeach (15, 16) adopts this second definition of rigidity and does not include extinction problems in his version of the water-jar test.

In the present investigation the *S*s were classified according to their performance on the water-jar test in such a way that either of these two criteria of rigidity could be employed. *S*s were classified as either flexible (F) or as manifesting one of three degrees of rigidity (R_1, R_2, or R_3) according to the following specifications:

F. *S* used the Einstellung formula for the Einstellung problems only.

R_1. *S* carried the Einstellung formula over to the first critical problem, but thenceforward did not use it.

R_2. *S* used the Einstellung formula for two or three of the critical problems but for none of the extinction problems.

R_3. *S* used the Einstellung formula in one or more of the extinction problems, offering it as his final solution, with no indication that he recognized that it did not provide a correct solution.

With this classification it is possible to use either criterion of rigidity:

1. *S*s classified as R_3 are rigid, and all others are considered flexible.

2. *S*s classified as R_1, R_2, and R_3 are rigid, and the three groups represent three degrees of rigidity.

Each hypothesis is subjected to two tests, corresponding to the two criteria of rigidity that have been proposed.

Criteria of Stress

Four groups of *S*s were tested for problem-solving rigidity under different conditions. Group A, the non-stress group, was tested under relaxed conditions with no time limit, and the three stress groups, Groups B, C, and D, received respectively one, two, and three increments of stress. The conditions of stress were intended to be mild and to approximate the degree of stress found in typical university examinations.

First increment. After giving the standard instructions for the water-jar test, *E* emphasized the importance of speed of performance and announced a time limit of ten minutes. At the end of three minutes he called out "only five minutes to go," and at the end of six minutes he said "three minutes left to go." He then remained silent until all *S*s had completed the test, and made no comment about the discrepancy between the announced time limit and the time it actually took for all *S*s to finish. (The longest time taken by any *S* was 22 minutes and the shortest 6 minutes.) Although the main stress was intended to be time pressure, the announcements introduced distraction, and a slight competitive note may also be involved in the group situation.

Second increment. After the speed instructions, *E* urged the *S*s to try to do well since the results of the test had been found to be related to academic success. This was intended to increase stress by increasing the degree of ego-involvement. In addition, a competitive component was introduced by instructing the *S*s to indicate completion of the test by raising the hand.

Third increment. A confederate was introduced as an expert in mathematics, who was interested in observing the *S*'s solution of problems. *S* was then given the test instructions and the special instructions for the first and second increments of stress. *E* and his confederate then took up positions on either side of *S*, looking over his shoulder while he worked.

The increments of stress were thus applied additively, without implying equality of the degrees of stress involved in each of the increments. Groups A, B, and C were tested in a group situation. The *S*s of Group D were tested individually.

Criterion of Insecurity

The criterion of insecurity was the total score derived from combining the scores on the insecurity subtests of four tests of security-insecurity constructed by Ainsworth and Ainsworth (2). In these tests, insecurity was given an entirely subjective definition in terms of feelings reported by *S* as characteristic of himself. The areas of adjustment covered by these four tests were familial relationships, extra-familial relationships, avocations, and philosophy of life. Each of the insecurity subtests included 12 items, making 48 insecurity items in all. A score of one was given for each of the items answered "true" and a score of ½ for each "?" answer. *S*s were classified into three subdivisions on the basis of their total insecurity scores.

"*Secure.*" Scores of 0 to 8.
"*Middle.*" Scores of 8½ to 14½.
"*Insecure.*" Scores of 15 to 32.

Criterion of Defensiveness

Each of the tests of security-insecurity devised by Ainsworth and Ainsworth (2) included a subtest of 12 items designed to measure the extent to which defense mechanisms were used in an attempt to reduce feelings of insecurity. A combined score was derived from these four subtests in the same way as the combined insecurity score was derived; this was the criterion of defensiveness employed in the present investigation. *S*s were classified into three subdivisions, as follows:

"*Low defensiveness.*" Scores of 0 to 12.
"*Moderate defensiveness.*" Scores of 12½ to 18.
"*High defensiveness.*" Scores of 18½ to 36.

Subjects

The research was undertaken in two phases. In Phase I, the tests of security-insecurity were administered to a total pool of 173 *S*s, first- and second-year students attending the University of London. From this pool 30 matched quartets were selected, the *S*s who composed each quartet belonging to the same insecurity subdivision and the same defensiveness subdivision. Thus, one quartet might be "secure" and of "low defensiveness," another "secure" and of "moderate defensiveness," and so on. One member of each quartet was then assigned to each of the four groups, A, B, C, and D. In Phase II of the experiment, the four groups were then tested for rigidity under four different sets of conditions, as described earlier.

As a check on matching procedures, group means were calculated, and the mean differences between groups were found for the measures of insecurity and defensiveness. All the mean differences were found to fall well below the .05 level of significance, and it was concluded that the groups were well matched. There were no significant differences between the groups in sex distribution, and although there were minor uncontrolled age differences, these were found to have no appreciable influence upon the results (1).

TABLE 1
INSECURITY AND RIGIDITY

Rigidity Subdivisions	Insecurity Subdivisions			Total
	Secure	Middle	Insecure	
F	9	5	1	15
R_1	7	7	4	18
R_2	12	18	6	36
R_3	8	18	25	51
Totals	36	48	36	120

Note.—1st test: $\chi^2 = 17.01$; $p < .01$; *df* 2 ($F + R_1 + R_2$ vs. R_3). 2nd test: $\chi^2 = 25.57$; $p < .01$; *df* 6.

RESULTS

Each hypothesis is subjected to two tests, corresponding to the two criteria of problem-solving rigidity that have been proposed. According to the first criterion, only *S*s classified as R_3 are considered to be rigid, and the test involves the comparison of R_3 *S*s with the rest of the sample. According to the second criterion, *S*s classified as R_1, R_2, and R_3 are all considered to be rigid and to represent three degrees of rigidity. The test employing this criterion of rigidity involves the whole distribution of *S*s as classified as F, R_1, R_2, or R_3.

Hypothesis 1: Insecurity and Rigidity

Table 1 shows the frequency of *S*s classified as F, R_1, R_2, and R_3 in each of the "secure," "middle," and "insecure" subdivisions. The first test of the hypothesis involves the comparison of *S*s classified as rigid (R_3) with the rest of the sample. It may be seen that this extreme form of rigidity tends to be characteristic of the insecure subdivision, and to occur less frequently in the middle subdivision, and still less frequently in the secure subdivision. The null hypothesis is rejected at the .01 level of significance.

The second test of the hypothesis involves the whole distribution. It may be seen that flexibility (F) and the least degree of rigidity (R_1) are relatively more frequent in the secure

TABLE 2
STRESS AND RIGIDITY
(N for each group is 30)

Rigidity Subdivisions	Control Group A	Stress Groups			Total
		Group B	Group C	Group D	
F	6	4	4	1	15
R_1	5	5	4	4	18
R_2	14	8	6	8	36
R_3	5	13	16	17	51

Note.—1st test: $\chi^2 = 10.12$; $p < .02$; df 3 ($F + R_1 + R_2$ vs. R_3). 2nd test: $\chi^2 = 14.31$; $p < .10$; df 9.

TABLE 3
DEFENSE MECHANISMS AND RIGIDITY

Rigidity Subdivisions	Defensiveness Subdivisions			Total
	Low	Moderate	High	
F	6	7	2	15
R_1	7	6	5	18
R_2	12	15	9	36
R_3	15	14	22	51
Totals	40	42	38	120

Note.—1st test: $\chi^2 = 5.52$; $p < .10$; df 2 ($F + R_1 + R_2$ vs. R_3). 2nd test: $\chi^2 = 6.99$; $p < .50$; df 6.

than in the middle subdivision, and more frequent in the middle than in the insecure subdivision. A moderate degree of rigidity (R_2) is more frequent in the middle subdivision than in either the secure or insecure subdivisions. And, as mentioned above, the most extreme form of rigidity (R_3) occurs most frequently in the insecure subdivision. The departure of this distribution from one that might be expected on the basis of chance is significant at the .01 level of significance.

The hypothesis is thus supported that rigidity in a problem-solving situation is related to insecurity in general life adjustment.

Hypothesis 2: Stress and Rigidity

Table 2 shows the frequency of *S*s classified as F, R_1, R_2, and R_3 in the nonstress group, Group A, and in each of the three stress groups, Groups B, C, and D. The first test of the hypothesis involves the comparison of *S*s classified as rigid (R_3) with the rest of the sample. It may be seen that rigidity thus defined is more frequent in the stress groups than in the nonstress group, and that rigidity increases in frequency with increasing stress. This tendency is significant at the .02 level of significance.

It is obvious that the greatest differences lie between the nonstress group, Group A, and the three stress groups, whereas the differences between the stress groups themselves are relatively small. Chi squares were calculated for each pair of groups separately, and Group A was found to differ significantly from each of the other groups with respect to R_3 behavior, but there were no other significant differences.

The second test, involving the whole distribution, does not support the hypothesis. Although there is a tendency for flexibility (F) and the milder forms of rigidity (R_1 and R_2) to decrease with increasing stress, as well as for the more extreme form of rigidity (R_3) to increase with increasing stress, the probability that this departs from a chance distribution is between .05 and .10, so that the null hypothesis cannot be rejected with confidence.

As a check on this finding, chi squares were calculated for each pair of groups separately, and the only significant departure from chance expectations was found when Group A was compared with Group D, the group receiving most stress.

Thus, only partial support for the second hypothesis is found. The experimental introduction of stress in the problem-solving situation increases the likelihood that the individual will behave rigidly therein, if rigidity is defined as the failure to shift when the problem requires a shift. The hypothesis is not supported if the definition of rigidity includes the milder forms in which a shift is not made when the problem permits but does not require a shift. Support is not found for that part of the hypothesis which states that rigidity will increase with increasing stress.

Hypothesis 3: Defensiveness and Rigidity

Table 3 shows the frequency of *S*s classified as F, R_1, R_2, and R_3 in each of the subdivisions of defensiveness—low, moderate, and high. The hypothesis that rigidity is not related to degree of defensiveness is supported by the first test, in which the *S*s classified as R_3 are compared with the rest of the sample. Although there is a tendency for R_3 behavior to occur more frequently among *S*s with high defensiveness than among those with moderate or low defensiveness, this tendency falls short of statistical significance.

When the hypothesis is submitted to the

second test, involving the whole distribution, it is clear that the null hypothesis cannot be rejected. According to both tests, the hypothesis is thus supported that rigidity in a problem-solving situation is unrelated to degree of defensiveness in general life adjustment.

Hypothesis 4: Insecurity, Stress, and Rigidity

In Fig. 1, the percentages of *S*s classified as F, R_1, R_2, and R_3 are shown for each of the "secure," "middle," and "insecure" subdivisions of the nonstress group and each of the three stress groups. Percentages are used instead of raw frequencies because there were more *S*s in the middle subdivisions than in the secure and insecure subdivisions.

In order to demonstrate that the independent variables of insecurity and stress act together in determining the incidence of rigidity it is necessary that (*a*) in each of the stress groups, Groups B, C, and D, there be a significant increase in the incidence of rigidity from the secure to the middle and from the middle to the insecure subdivision, and that (*b*) in each of the secure, middle, and insecure subdivisions there be a significant increase in the incidence of rigidity from Group B to Group C, and from Group C to Group D.

In the first test of the hypothesis the relative frequency of R_3 behavior in the various subdivisions of the four groups of *S*s is of concern. In Group A, the nonstress group, the only *S*s who are classed as R_3 are insecure *S*s. In each of the three stress groups, *S*s classed as R_3 occur in the secure and middle subdivisions as well as in the insecure subdivisions. In each of these stress groups more middle than secure *S*s and more insecure than middle *S*s are classed as R_3. In each case the findings are thus exactly in accordance with expectations. Only once in 216 times, i.e., $(1/6)^3$ could this predicted order of subdivisions, (secure < middle < insecure), be expected to occur by chance in three independent samples. Thus the first half of the hypothesis is supported at a highly significant level; the more insecure the person, the more likely he will be to behave rigidly under stress. This effect is consistent at all three levels of stress represented in this study.

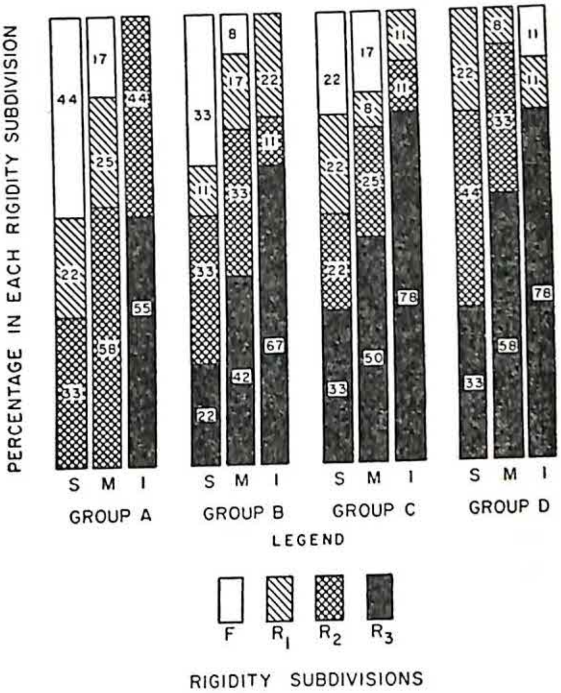


FIG. 1. PERCENTAGES OF THE SECURE, MIDDLE, AND INSECURE SUBDIVISIONS OF EACH OF THE FOUR STRESS GROUPS SHOWING EACH OF FOUR DEGREES OF RIGIDITY

The second half of the interaction hypothesis provides that the greater the stress, the greater the incidence of rigidity. The hypothesis holds good when Group B is compared with Group C. In the case of each of the secure, middle, and insecure subdivisions, the incidence of rigidity is greater in Group C than in Group B. But when Group C is compared with Group D, expectations are met only in the case of the middle subdivision, in which Group D is more rigid than Group C. The incidence of rigidity in Group C and Group D is equal for both of the secure and insecure subdivisions. Therefore, the second half of the hypothesis is not supported, since there is no significant tendency for incidence of rigidity to increase with increased stress.

In the second test of the hypothesis, all degrees of rigidity (R_1, R_2, and R_3) are involved, in comparison with each other and with flexibility (F). It may be seen that in both Group B and Group C the secure subdivision is more flexible than the middle subdivision, and the middle subdivision is more flexible than the insecure subdivision. This is the case both when one considers the *S*s classified as F alone, and when F and R_1 are combined. But the same tendency does not

---

[3] These three *S*s were discarded from the sample and their places taken in their respective quartets by three other *S*s from the pool.

hold for Group D. Thus the first half of the hypothesis is not supported. When F and R_1 incidence is examined within each of the secure, middle, and insecure subdivisions, it may be seen that there is no regular tendency for flexibility to decrease with increasing stress. Thus the second test fails to support the hypothesis.

DISCUSSION

This investigation demonstrates that the extent to which situational influences affect rigidity depends upon the criterion of rigidity employed. The first criterion used in this study follows Luchins (11, 12, 13, 14) in classing as rigid only those *S*s who fail to shift to a new method of solution of the problem when the situation demands a shift. When this criterion is used, support is found for Luchins' claim that situational factors influence the occurrence of rigidity. The introduction of experimental stress is found significantly to increase the failure to shift when the situation demands a shift.

The second criterion distinguishes degrees of rigidity in inverse proportion to the readiness with which a shift to a new method of solution is made when the situation permits but does not demand a shift. This criterion approximates that used by Rokeach (15, 16). When the milder degrees of rigidity are taken into consideration, the influence of situational factors upon rigidity are obscured. The introduction of experimental stress does not significantly decrease the readiness to shift before the situation demands a shift.

It is suggested that some of the apparent contradictions in the literature of the water-jar test may be attributable to differences in the criteria of rigidity employed. Until a consensus is reached regarding the basic conditions that influence the incidence of problem-solving rigidity, it would seem wise to adhere to the more extreme criterion, namely, failure to shift when the situation demands a shift. It is clear that this behavior may be described as rigid. It is not clear that it is rigid to use a relatively complicated solution instead of a relatively direct solution to a problem, when both solutions are correct, and when the more complicated solution may well be the more economical because the preceding problems have made it the solution most readily available to the person.

When rigidity is defined as failure to shift when the situation demands a shift, it is found to be influenced both by situational factors and by the degree of insecurity experienced by the person in his general life adjustment. Indeed, under the conditions of the present experiment, insecurity characteristic of the person seems to play a larger part in determining rigid behavior than does insecurity induced by situational manipulation.

Under the conditions of the present experiment, the incidence of rigidity is not found to increase significantly with an increase in stress, although there are significant differences in incidence of rigidity in stress situations in contrast with the nonstress situation. It seems probable that a closer correspondence between increasing stress and increasing incidence of rigidity could be demonstrated in an experiment in which the stress increments were larger. It seems a reasonable hypothesis that the introduction of severe stress might override or cloak the contribution made by personality-related insecurity. This hypothesis implies that there is a limit beyond which even the relatively secure individual can not maintain flexibility under stress. In the present investigation this limit is not approached, since only a third of the secure *S*s behave rigidly under the most severe stress applied here.

It is relevant to the discussion of limits of stress-tolerance to draw attention to the disorganization that occurs with a few *S*s, a finding to which reference has not heretofore been made. In some of the problems of the water-jar test, ten *S*s neither used the Einstellung formula nor the simple, direct solution; instead, they attempted long and complicated solutions that were arithmetically incorrect, and further betrayed disorganization by other signs such as scratching out false starts. Seven of the ten were classified as R_3, and nine were in the stress groups. They were distributed irregularly through the security-insecurity subdivisions. Furthermore, there were three other *S*s who were discarded[4] because their water-jar test performances were unscorable according to the criteria used. Two of these three *S*s were clearly disorganized. These were tested under Group D conditions where their behavior could be observed. They trembled and seemed completely blocked and close to tears. (The third *S* may also have been disorganized to the point of complete blocking for he attempted none of the problems; this *S* was tested under group conditions, so it was not possible to ascertain the reason for his failure.)

---

[4] See Footnote 3.

The phenomenon of disorganization has not heretofore been reported in the literature of the water-jar test. This seems strange, since the literature on stress reports, as the most immediate and striking effect of stress, a disorganized and diffuse pattern of response (e.g., 3, 9, 17). It is suggested that two practices with respect to the water-jar test have tended to obscure the emergence of disorganization: (*a*) the general practice of administering the test in a group situation in which the behavior of the *S* can be observed only through his pencil-paper record, and (*b*) the general practice of discarding from the data test records that include arithmetical errors other than those attributable to the Einstellung effect, thus discarding the evidence for disorganization.

It is suggested that, to the extent that the individual experiences insecurity in the problem-solving situation, this insecurity manifests itself in tendencies towards a diffuse and disorganized response, even though these tendencies are not easily observable in the test record under group test conditions. The Einstellung formula offers a relatively secure alternative to disorganization, and hence the disorganized *S*s tend to return to it, even though in so doing they avoid tackling the solution to the problem. In this way, rigidity may be viewed as a defense against insecurity experienced in the problem-solving situation. Thus it is believed that rigidity may emerge as a defense in a specific situation even though it is not found to be related significantly to the degree of defensiveness in general life adjustment.

It is suggested that the positive relationship between rigidity and general defensiveness reported frequently, but by no means unanimously, in the literature may be accounted for by the fact that highly defensive persons tend to be relatively insecure. This was clearly demonstrated in the case of the measures used in the present study by Ainsworth and Ainsworth (2) who found significant positive relationships between scores on all the insecurity subtests of the security-insecurity tests and the subtests measuring degree of defensiveness. The hypothesis consistent with this writer's position is that to the extent that defense mechanisms are effective in protecting the individual against feelings of insecurity, he should be able to meet a relatively nonstressful problem-solving situation without appreciable insecurity and deal flexibly with it. On the other hand, if the person is relatively insecure despite considerable use of defense mechanisms—and this is more usually the case—he would be rigid in a problem-solving situation. These hypotheses are supported in the present investigation, since rigidity is found to be significantly related to insecurity but not to degree of defensiveness.

SUMMARY

1. The general hypothesis of the study is that problem-solving rigidity is attributable to insecurity experienced in the problem-solving situation. This hypothesis is believed to offer a resolution of the controversy as to whether rigidity is a situational phenomenon or attributable to a generalized personality characteristic. Insecurity is the bridge-concept between these two apparently contradictory views, since the degree of insecurity experienced in a specific situation is held to be influenced both by the stressfulness of the situation itself and by the degree of insecurity experienced by the individual in his general life adjustment.

2. Four specific hypotheses were formulated and may be summarized as follows: (*a*) rigidity is related to insecurity in general life adjustment; (*b*) rigidity is related to stress; (*c*) rigidity is not related to degree of defensiveness in general life adjustment; and (*d*) the introduction of stress tends to act differentially in producing rigidity in accordance with the degree of insecurity experienced in general life adjustment.

3. *S*s were 120 university students divided into four groups of 30 each, matched on the

basis of insecurity and defensiveness in general life adjustment. A water-jar test of problem-solving rigidity was administered to each group under different conditions, one group serving as a control group and the other three groups receiving respectively one, two, and three increments of stress. Two alternative criteria of rigidity were employed. The first criterion defined rigidity as a failure to shift to a new method of solution when the problem requires a shift. The second criterion distinguished two milder degrees of ridigity as well as the more extreme form of rigidity implied by the first criterion.

4. Significant support is found for the hypothesis that rigidity is related to insecurity in general life adjustment. Support is also found for the hypothesis that rigidity is related to situational stress, but only when rigidity is defined as a failure to shift when the problem requires a shift, and not when milder degrees of rigidity are also considered. Although the incidence of rigidity is significantly greater in the stress groups than in the nonstress groups, the incidence does not increase significantly as increments of stress are added. Because of this latter finding it is not possible to demonstrate interaction between the two variables of stress and insecurity, but the relationship between rigidity and degree of insecurity in general life adjustment is found to hold under three different degrees of stress. Support is found for the hypothesis that rigidity is not significantly related to degree of defensiveness in general life adjustment.

5. The implications of these findings are discussed, with particular stress upon the influence of the criterion of rigidity employed. Attention is drawn to the phenomenon of disorganization that occurred with a few *Ss*, and the hypothesis is formulated that insecurity in the problem-solving situation manifests itself in tendencies towards a diffuse and disorganized response, and that it is against these tendencies that rigidity acts as a defense. This hypothesis does not imply a generalized rigidity factor characteristic of the personality.

REFERENCES

1. Ainsworth, L. H. A study of rigidity. Unpublished doctoral thesis, University College, Univer. of London, 1953.
2. Ainsworth, Mary D., & Ainsworth, L. H. *Measuring security and insecurity.* Toronto: Univer. of Toronto Press, in press.
3. Barker, R. G., Dembo, T., & Lewin, K. Frustration and regression: An experiment with young children. *Univer. Ia. Stud. in Child Welf.*, 1941, **18,** No. 1.
4. Brown, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48,** 469–476.
5. Cattell, R. B. The riddle of perseveration. II. Solution in terms of personality structure. *J. Pers.* 1946, **14,** 239–267.
6. Cattell, R. B., & Tiner, G. The varieties of structural rigidity. *J. Pers.*, 1949, **17,** 321–341.
7. Eriksen, S. W., & Eisenstein, D. Personality rigidity and the Rorschach. *J. Pers.*, 1952–53, **21,** 386–391.
8. Frenkel-Brunswik, Else. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18,** 108–143.
9. Goldstein, K. *The organism.* New York: American Book, 1939.
10. Levitt, E. E. The water-jar Einstellung test as a measure of rigidity. *Psychol. Bull.*, 1956, **53,** 347–370.
11. Luchins, A. S. Mechanization in problem-solving. *Psychol. Monogr.*, 1942, **54,** No. 6.
12. Luchins, A. S. Rigidity and ethnocentrism: a critique. *J. Pers.*, 1949, **17,** 449–466.
13. Luchins, A. S. On recent usage of the Einstellung effect as a test of rigidity. *J. consult. Psychol.*, 1951, **15,** 89–94.
14. Luchins, A. S. The Einstellung test of rigidity: its relation to concreteness of thinking. *J. consult. Psychol.*, 1951, **15,** 303–310.
15. Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43,** 259–278.
16. Rokeach, M. Prejudice, concreteness of thinking and reification of thinking. *J. abnorm. soc. Psychol.*, 1951, **46,** 83–91.
17. Selye, H. *The physiology and pathology of exposure to stress.* Montreal: Acta, 1950.
18. Wesley, Elizabeth. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1954, **48,** 129–134.

Received April 12, 1957.

INTELLIGENCE AND EMOTIONAL DISTURBANCE: EVIDENCE FROM ARMY AND VETERANS ADMINISTRATION RECORDS[1]

JOHN B. MINER[2] AND JAMES K. ANDERSON

Conservation of Human Resources Project, Columbia University

For many years, psychologists have been attempting to determine whether or not there is a difference on the average in intellectual level between individuals suffering from emotional disturbance and the normal population. The first studies appeared during World War I, and interest continued into the 1920's. Then, after a hiatus in the late twenties and early thirties, a flood of publications on the subject appeared as large numbers of psychologists began to accept positions on mental hospital staffs. By 1945, however, this activity had virtually ceased. Since World War II only a handful of studies has appeared, most of which have been tangential by-products of more broadly conceived investigations.

The decline in research activity seems to have resulted not from a lack of interest, but from an increased understanding of certain basic methodological difficulties. The results of the earlier studies were far from conclusive. In some instances, lower intelligence was found throughout all the emotionally disturbed groups studied. Other investigations reported lower test scores only among certain diagnostic groups. Still others failed to establish any differences, and, on occasion, pathological groups were well above the general level of the population in intelligence. It became apparent that a definitive answer cannot be obtained by testing a group of patients in a single hospital and comparing their scores with a control group whose representativeness is equally questionable.

However, sample selection was not the only factor contributing to the inconclusive results. As Roe and Shakow (8) demonstrated, the test scores of psychotics and even neurotics are frequently lowered by a variety of distractors that are much more closely associated with the psychiatric condition itself than with the level of intellectual functioning. Such distractors include those extraneous thoughts, feelings, and perceptions that intrude upon consciousness and thus interfere with a total concentration on the test items. Thus, hallucinations, obsessive ruminations, anxiety attacks, and flights into fantasy frequently occur in the test situation. Accordingly, performance is often spotty with easy items failed and difficult ones passed. The scores on timed tests are particularly likely to be lowered by these distractors. Certainly estimates of intellectual level obtained under such conditions cannot be considered as valid indicators of optimum or even typical functioning. Furthermore, if patients must be tested during a period of emotional illness, as when the problem of intellectual deterioration is under investigation, controls must be provided that will permit a clear differentiation between the effects of distractors and any possible deterioration effects.

Although the results of previous studies fail to establish a lower level of intelligence as characteristic of the emotionally disturbed, certain facts suggest that this is a tenable hypothesis. Since 1949, the staff of the Conservation of Human Resources Project has been studying military manpower utilization during World War II. This work has already resulted in the publication of two books, *The Uneducated* (4) and *Psychiatry and Military Manpower Policy* (5). Three additional volumes dealing with ineffective performance as related to the military and post-war civilian environments are now in preparation. In the course of these investigations it has become evident that those servicemen who were discharged for psychosis and neurosis were as a group well below the army as a whole in educational attainment at the time of induction. Comparative data are presented in Table 1. It is also clear that there is a high positive relationship between education and intelligence. This conclusion has been substantiated elsewhere (7) employing evidence derived from a representative sample of

---

[1] Drs. Eli Ginzberg Director of the project, Sol W. Ginsburg consultant in psychiatry, and John L. Herma staff psychologist reviewed this paper and made valuable suggestions. Similar contributions were made by Drs. Irving Lorge of Teachers College, Columbia University, Silvan S. Tomkins of Princeton University, and Douglas W. Bray of the American Telephone and Telegraph Company.

[2] Now with the Industrial Relations Department, Atlantic Refining Co., Philadelphia, Pa.

TABLE 1

PER CENT OF ARMY NEUROTICS, PSYCHOTICS, AND TOTAL INDUCTEES ATTAINING VARIOUS EDUCATIONAL LEVELS

Group	School Years Completed			
	0–8	9–11	12	13+
Neurotics	41	31	20	8
Psychotics	40	30	20	10
Inductees	32	28	26	14

the United States population. Other investigators have obtained correlation coefficients varying generally in the range .50 to .75 using more restricted samples. The median value appears to be approximately .65. These findings suggest the following hypothesis:

1. The intelligence of neurotics and psychotics prior to the time of breakdown is below that of the population as a whole.

In the belief that intelligence may in many cases be an important factor in the resolution of emotional conflicts at the conscious level and thus in a recovery, a second hypothesis was formulated.

2. The pre-breakdown intelligence of individuals who later develop a severe psychiatric disorder of a *chronic* nature is below that of individuals who do not *remain* severely disturbed.

METHOD

In the investigation of these hypotheses, materials collected in connection with the Conservation of Human Resources Project's study of ineffective performance during World War II were employed. The cases derive from a special sample of World War II enlisted men retained by the Adjutant General's Office in Washington. From this sample, which consisted of five per cent of all enlisted separations from the army, all individuals inducted during the last four months of 1942 who had been prematurely discharged for inaptness, undesirability, psychosis, and psychoneurosis were selected. From this prematurely discharged group, special samples were selected to be representative within the following categories:

Service in the U. S. of less than six months.
Service in the U. S. of six to 18 months.
Service in the U. S. of 18 months or more.
Service overseas in the ground forces.
Service overseas in the Air Force.
Service overseas in the service forces.

There were 231 psychoneurotics and 120 psychotics included. By weighting the data within categories, it was possible to obtain figures representative of enlisted men inducted during the last four months of 1942. Approximately one sixth of the over ten million army accessions in World War II occurred during this period.

Extensive information on the subjects (*S*s) was obtained from the files of the Army Records Center in St. Louis. These files contain service records, medical records, and separation board proceedings. Follow-up information has also been made available from the files of the Veterans Administration. Many of the *S*s obtained disability compensation and medical treatment from the Veterans Administration. Others received extensive education and training, frequently coupled with vocational guidance interviews and testing. These contacts have resulted in the accumulation of a variety of records that in most cases give a comprehensive picture of the veterans' adjustment since discharge. When the Veterans Administration data were, for some reason, incomplete, they were supplemented through the use of questionnaires mailed to the veterans' homes. In general, however, since nearly all soldiers discharged for psychosis and many psychoneurotics received disability compensation for a number of years and were accordingly examined at frequent intervals, the VA records on these groups are quite complete as checked through 1953.

Because many of the Form 20s, which listed Army General Classification Test scores for each soldier, were destroyed after the War, it was impossible to obtain information on the intelligence of all *S*s. Scores for tests administered at the time of induction were available on 103 (44.6 per cent) of the neurotics and 61 (50.8 per cent) of the psychotics. These samples could have been supplemented if intelligence test scores listed in the medical records and obtained during a period of hospitalization for emotional disturbance had been included. However, this would have contaminated the results by introducing an unknown number of scores that had been artificially affected by disease-related distractors. Although most of the AGCT results were obtained from army records, there were some instances in which they were available only in the VA files. It is, of course, well known that testing conditions at the various induction centers were not completely standardized. Nevertheless, the error thus introduced does not appear to be sufficiently large or sufficiently systematic to invalidate any major conclusions.

The diagnostic composition of the sample is indicated in Table 2. These diagnoses were given at the time of discharge from service and were obtained from the medical records. Because of the lack of any standard method of classification that could be applied throughout World War II and the very sizable differences in training and experience among the psychiatrists making the diagnoses, these data can be considered as suggestive only. The major division into psychoneurotic and psychotic groups is, however, in general quite consistent with the impressions derived from a thorough reading of the medical histories by several Conservation of Human Resources Project staff members and appears to be reliable. It should be emphasized that these samples contain only those men who were discharged from the Army and who received a psychiatric diagnosis. This diagnosis in most cases reflected a severe disturbance in the ability to perform. However, there were some men who, although they did suffer from symptoms of an emotional origin, were discharged only because their symptoms came to the attention of the

TABLE 2

DISCHARGE DIAGNOSES IN NEUROTIC AND PSYCHOTIC SAMPLES

Diagnosis	*N*	%
Psychoneurosis		
Anxiety state	54	52.4
Mixed type	13	12.6
Conversion hysteria	11	10.7
Hysteria	7	6.8
Neurocirculatory asthenia	5	4.9
Neurasthenia	4	3.9
Reactive depression	4	3.9
Obsessive-compulsive	2	1.9
Unspecified type	2	1.9
Hypochondriasis	1	1.0
Psychosis		
Dementia praecox, paranoid	14	23.0
Dementia praecox, unspecified	11	18.0
Unspecified type	8	13.1
Dementia praecox, simple	6	9.8
Dementia praecox, catatonic	4	6.6
With Mental Deficiency	4	6.6
Schizo-affective type	3	4.9
Dementia praecox, hebephrenic	3	4.9
Dementia praecox, mixed	2	3.3
Manic-depressive, manic	2	3.3
With Psychopathic Personality	2	3.3
Manic-depressive, depressed	1	1.6
With Epilepsy	1	1.6

authorities during a period when the army's personnel policy made separation relatively easy. These individuals, whose performance was not grossly affected by their symptoms, were more likely to be discharged for psychoneurosis than for psychosis.

Information on the AGCT scores for total inductions in the period September, 1942, through May, 1943, was obtained from Davenport (3). Because the neurotic and psychotic samples contained only enlisted men these data were corrected to exclude men who later became army officers. This control group differs from the experimental samples in that it includes men inducted during the first five months of 1943. There is no reason, however, to believe that the early 1943 inductees differed in AGCT score from those who entered the army during late 1942. It was not possible to eliminate from the control group those men who were subsequently discharged for psychiatric reasons.

In order to investigate the first hypothesis, comparisons were made between the distributions of AGCT scores for the neurotics, psychotics, and controls. The distributions for the two pathological groups were obtained by weighting the data within each of the six length-of-service and branch-of-service categories according to the incidence of neurosis and psychosis in each category. The weighting factors were derived from the five per cent sample retained by the Adjutant General's Office. Significance was tested by the chi square method because control data were available only in terms of gross score ranges.

As previously indicated, any definitive test of this hypothesis must employ an experimental design that rules out the effects of biased sampling and disease-related distractors as well as any possible deterioration effects. In the use of military data, a major sampling difficulty arises because AGCT scores are not available for all *S*s. As a check on the possibility of sample selection, comparisons on education, race, and age were made between the subsample whose AGCT scores were available and a group whose scores were lost. Since all of these variables have a sizable and, to some degree, independent relationship with intelligence (7), failure to establish differences between the two groups would seem to offer strong evidence in support of the hypothesis that selection on intelligence has not occurred. In order to investigate the possibility that emotional disturbance existing at induction may have lowered test scores, the AGCT distribution of those who broke down during the first three months of service was compared with the distribution for those whose illness appeared later. Breakdown was defined as any contact with the medical service, whether or not hospitalization was required, that resulted in a psychiatric diagnosis. The assumption was that if those who suffered from severe mental disorders were not diagnosed and rejected at the time of the induction examination, their condition would, nevertheless, in most instances become evident within the first three months of service. Failure to establish a difference in AGCT score between the early and late breakdown groups would indicate either that illness at induction did not have an effect on the intelligence test results of this group or that even those who broke down early were free of disease at the time of testing. It is, of course, true that failure to establish a difference might result if the early breakdown group had a higher pre-illness intelligence than the late breakdown group but was disturbed at the time of induction testing and thus reflected lowered scores. Such a situation seems highly improbable. In any event, a finding of no difference in intelligence between the early and late breakdown groups would seem to offer strong presumptive evidence in support of the conclusion that any differences found between the control and pathological groups could not be attributed to the effects of disease-related distractors or deterioration.

Finally, in order to investigate the second hypothesis, the AGCT distribution of those with a post-war history of hospitalization for emotional disorder was compared with the distribution for those with no such history. Postwar is understood to include the period from discharge through 1953. The data on hospitalization came almost exclusively from the VA records, although the questionnaires were also inspected. With the exception of three cases, all members of the hospitalized sample were at one time or another in VA facilities. In some cases, institutionalization followed immediately on discharge, whereas in others, a period of freedom from hospital care intervened. Multiple admissions were not uncommon. Since a number of men discharged from the service for psychoneurosis were later hospitalized for psychosis and a few discharged for psychosis later received a diagnosis of psychoneurosis, no attempt was made to differentiate between these two groups. Of the 64 *S*s hospitalized after the war, 16 were considered psychoneurotic at the time of separation from service and 48 psychotic. In view of the VA's continual interest in the psychiatric status of these men and the fact that the men themselves stood to benefit through increased pension payments by keeping the VA informed of any hospitalization for emotional disturbance, there is

TABLE 3

PER CENT OF NEUROTICS, PSYCHOTICS, AND CONTROLS SCORING AT VARIOUS AGCT LEVELS

Group	N	AGCT Scores			
		−59	60–89	90–109	110+
Neurotics[a]	103	15.8	42.5	22.0	19.7
Psychotics[a]	61	22.2	33.2	32.4	12.2
Controls		10.6	31.8	33.0	24.6

[a] Based on weighted data.

little reason to believe the data on institutionalization are significantly in error.

RESULTS

The weighted percentage distribution for the neurotics and the psychotics and the percentage distribution for the controls are presented in Table 3. The standard grouping of AGCT scores was used except that Groups I and II (scores of over 129 and 110 through 129 respectively) were combined because of the small number of cases in each of these two categories among the psychiatric samples. The control *S*s differ significantly from both the neurotics and the psychotics. (The neurotic-control comparison yields a χ^2 of 11.16 with $p < .02$, while the psychotic-control comparison yields a χ^2 of 11.59 with $p < .01$.) It is clear that, in accordance with the first hypothesis, both psychiatric groups score well below the control sample. In addition, the neurotics and psychotics do not differ reliably from one another with respect to AGCT score (χ^2 of 4.55, $p > .10$). The mean weighted AGCT score for the neurotics is 81.2 and for the psychotics 84.7. Because control data are available only in terms of gross score ranges and because the control sample includes men who were later discharged for psychiatric reasons, it is difficult to specify a precise figure for the normal group. However, the mean is almost certainly not below 94 and may well be several points higher.

Data on the educational level attained by the *S*s whose AGCT scores were available and those for whom such scores were not available are presented in Table 4. When both psychiatric groups are combined, it is clear that those whose AGCT scores are available have a higher educational level (χ^2 of 8.46, $p < .05$). Within both the neurotic and psychotic groups the educational difference between those with and those without AGCT scores is in the same direction but does not attain significance (p values between .10 and .20 in both cases). Since, in addition, educational level and AGCT score have a high relationship with each other in both groups ($r = .74$ for the neurotics and .60 for the psychotics), it seems probable that the mean scores for the psychiatric groups are, if anything, too high. The educational selection of the particular samples used in this study appears to have resulted in an underestimate of the intelligence differential between those suffering from emotional disturbance and normals. It is possible that a similar selection factor was operative in other studies employing World War II data (2, 6) which yielded results somewhat at variance with those obtained here. However, without cross-validation the existence of a stable selective bias cannot be definitely established.

TABLE 4

PER CENT OF AGCT AND NON-AGCT SAMPLES ATTAINING VARIOUS EDUCATIONAL LEVELS

Sample	N	School Years Completed			
		0–6	7–8	9–11	12+
AGCT	164	17.7	18.3	31.1	32.9
Non-AGCT	187	26.2	24.6	26.7	22.5

TABLE 5

PER CENT OF AGCT AND NON-AGCT SAMPLES OF WHITE AND NEGRO RACE

Sample	N	Race	
		White	Negro
AGCT	164	81.7	18.3
Non-AGCT	187	85.0	15.0

As indicated in Table 5, the AGCT and non-AGCT samples differ little in racial composition (χ^2 for the total psychiatric group is .48, for the neurotics .01, and for the psychotics 1.68, all with $p > .10$). Similarly, the two samples do not differ with respect to age. (For the total psychiatric group shown in Table 6, the χ^2 is 3.46, for the neurotics alone 2.60, and for the psychotics alone 5.04, all with $p > .10$.)

Table 7 presents the AGCT distributions for those men who exhibited psychiatric symptoms during their first three months of service and for those who did not exhibit symptoms until later. The two groups do not differ significantly (χ^2 is .07, $p > .10$). The unweighted

TABLE 6

PER CENT OF AGCT AND NON-AGCT SAMPLES OF VARIOUS AGES AT INDUCTION

Sample	N	Years			
		18–24	25–29	30–34	35+
AGCT	164	64.0	14.0	12.2	9.8
Non-AGCT	187	65.2	19.3	8.6	6.9

TABLE 7

PER CENT OF EARLY AND LATE BREAKDOWN GROUPS SCORING AT VARIOUS AGCT LEVELS

Group	N	AGCT Scores			
		−59	60–89	90–109	110+
Early breakdown	20	15.0	35.0	30.0	20.0
Late breakdown	144	14.0	33.3	32.6	20.1

TABLE 8

PER CENT OF THOSE HOSPITALIZED AND NOT HOSPITALIZED AFTER DISCHARGE SCORING AT VARIOUS AGCT LEVELS

Group	N	AGCT Scores			
		−59	60–89	90–109	110+
Hospitalized	64	26.6	35.9	28.1	9.4
Not hospitalized	100	6.0	32.0	35.0	27.0

mean for the early breakdown group is 87.5; for the later it is 88.4. These results strongly suggest that neither intellectual deterioration nor illness-related distractors affected test performance in such a way as to produce the differential noted between the AGCT scores of the psychiatric groups and the controls.

Evidence on the intelligence of those who were hospitalized for psychiatric disorders after separation from the Army as compared with the nonhospitalized group is presented in Table 8. There is a clear-cut tendency to lower scores among the hospitalized group (χ^2 is 17.38, $p < .01$). The mean AGCT score of the hospitalized group is 78.3 and of the nonhospitalized 95.8. This result supports the second hypothesis.

DISCUSSION

Strictly speaking, these findings can only be generalized to that portion of the World War II army that was inducted during the last four months of 1942. Since, however, this includes approximately one sixth of those who served, and since there is no reason to believe that this group differed in any respect relevant to the results from the army as a whole, the conclusions can, with some degree of confidence, be applied to the total group who served in World War II. It seems probable that similar results would be obtained in a study based on the population as a whole. However, due to the restricted age range of the military group, as well as its higher educational level, this generalization cannot be made with certainty. Furthermore, it cannot be assumed without further study that identical results would be obtained if a different set of operations for defining intelligence, e.g., a test other than the AGCT, was employed.

Although the fact of a lower level of intelligence among the emotionally disturbed groups has been established, the findings offer little in the way of clues to an understanding of why this is the case. Theoretically, there are three possibilities. First, limited intelligence may, under certain circumstances, contribute to the development of a psychiatric disorder. Second, character disturbances existing at an early age and later eventuating in a true emotional disorder may act to inhibit learning and the development of native potential. Third, some other factor or factors may operate to produce both a lowered level of intelligence and a predisposition to breakdown.

Although further study is certainly required, it seems probable that all three factors are operative. The case materials collected in connection with the Conservation of Human Resources Project studies of ineffective performance are particularly revealing with regard to the hypothesis that limited intelligence may contribute to the development of a psychiatric disorder. Although many men were unable to perform adequately in the military situation because of marked vulnerability under stress (low ego strength or emotional instability), and many others were severely lacking in the necessary intelligence, there was, in addition, a third group in which these two factors seemed to overlap. These men apparently failed because in the military setting they were exposed for the first time to demands for learning and understanding beyond their intellectual competence. They became threatened, anxious, and eventually their performance was severely affected. It seems likely that the lower the level of a man's intelligence, the greater is the probability that, at some time in his life, he will be exposed to environmental demands with which he is unable to cope. If these demands become so insistent as to make him acutely aware of his intellectual deficiencies, intense feelings of inadequacy may be precipitated and a severe emotional disturbance ensue.

It also seems probable that some types of character disturbances which are of insufficient severity to be considered true emotional disorders may operate from an early age to inhibit the development of innate learning ability. Evidence in support of this conclusion has been presented by Weisskopf (10) and by Sontag, Baker, and Nelson (9). Motivation to learn may not develop or may be misdirected so that the knowledge that our society values and thus includes in the school curriculum and intelligence tests is not absorbed. Since, in such individuals, ego strength may be assumed to be minimal and susceptibility to stress high, the probability of emotional breakdown is well above that of the population as a whole. Thus, early character disturbances associated with low ego strength would seem not only to predispose to later severe maladjustment but also to limit the development of intellectual capacities.

The third notion predicates such factors as a generally deprived environment as producing a lower intelligence among the emotionally ill. Elsewhere (7) it has been indicated that the level of intellectual functioning is a product of the interaction of three variables: native potential, motivation to learn, and environmental stimulus potential. It seems highly probable that a child who is brought up in an environment in which motivation to learn is not instilled, or in which environmental stimulus potential is minimal, or whose parents are of limited innate intelligence, also may well grow up in an environment that predisposes him to a psychiatric condition. In other words, intellectual deprivation is probably frequently associated with emotional deprivation with the same individual subjected to both. It is self-evident that this need not always be the case and that emotional deprivation may occur even under conditions that lead to the greatest development of native intellectual potential. Although it has been assumed that this association occurs as a function of the nature of the environment, the possibility of a hereditary linkage cannot be ruled out. It may be that a genetic factor that limits the development of intelligence frequently occurs in conjunction with a hereditary predisposition to emotional disturbance.

The findings also raise certain questions with regard to intellectual deterioration in the emotionally ill. If those who suffer from functional disorders are on the average of lower intelligence than the population as a whole, and if those who remain ill or have a recurrence are as a group even less intelligent than those who recover relatively quickly, initial intelligence must be controlled in any study aimed at determining whether deterioration actually occurs. Without such control it is impossible to say whether any obtained difference is primarily attributable to a lower level of initial intelligence in the chronic group or basically a reflection of true deterioration. Recent studies (1, 7) have raised serious doubts as to the existence of intellectual deterioration among the older age groups of the population. A reexamination of the deterioration concept as applied to those suffering from chronic emotional disturbance also seems in order.

SUMMARY

AGCT scores at the time of entry into military service were obtained for representative samples of men inducted during late 1942 and prematurely separated for psychosis and psychoneurosis. A comparison of these scores with those for total inductions in the period September 1942 through May 1943 revealed a reliably lower intelligence among both psychiatric groups. However, no reliable difference was found between the neurotics and psychotics. A further comparison between the scores of those prematurely discharged soldiers who were hospitalized for emotional disturbance as civilians and the scores of those not hospitalized indicated that the hospitalized veterans originally had a much lower intelligence level.

REFERENCES

1. BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, **10**, 805–818.
2. BRILL, N. Q., & BEEBE, G. W. *A follow-up study of war neuroses.* Washington: U. S. Veterans Administration, 1955.
3. DAVENPORT, R. K. Implication of military selection and classification in relation to universal military training. *J. Negro Educ.*, 1946, **15**, 585–594.

4. GINZBERG, E., & BRAY, D. W. *The uneducated.* New York: Columbia Univer. Press, 1953.
5. GINZBERG, E., HERMA, J. L., & GINSBURG, S. W. *Psychiatry and military manpower policy.* New York: Columbia Univer. Press, 1953.
6. MASON, C. F. Pre-illness intelligence of mental hospital patients. *J. consult. Psychol.*, 1956, **20,** 297–300.
7. MINER, J. B. *Intelligence in the United States.* New York: Springer, 1957.
8. ROE, ANNE, & SHAKOW, D. Intelligence in mental disorder. *Ann. N.Y. Acad. Sci.*, 1942, **42,** 361–490.
9. SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Personality as a determinant of performance. *Amer. J. Orthopsychiat.*, 1955, **25,** 555–562.
10. WEISSKOPF, EDITH A. Intellectual malfunctioning and personality. *J. abnorm. soc. Psychol.*, 1951, **46,** 410–423.

Received April 10, 1957.

INFLUENCING ETHNOCENTRISM IN SMALL DISCUSSION GROUPS THROUGH A FILM COMMUNICATION[1]

LEONARD L. MITNICK[2] AND ELLIOTT McGINNIES

University of Maryland

THE experimental literature dealing with persuasive communications has been summarized and given a theoretical orientation by Hovland et al. (6). Many questions, however, remain unanswered, among them problems concerning the relationships between certain aspects of group composition, interaction, and the direction and stability of attitude change under the impact of a persuasive communication. The present paper is concerned with learning and attitude change among the members of small, attitudinally homogeneous groups that experienced varying degrees of participation with respect to discussion of a motion picture film dealing with the problem of group prejudice and conflict.

Motion picture films may exert a significant influence upon the attitudes and opinions of viewers (7, 14, 15, 16). It has also been shown that group participation in decision-making is a more effective technique for generating acceptance than mere exposure to persuasive influences (1, 2, 10). Active participation either imposed by the experimenter or allowed to develop at the initiative of the group members seems to encourage acceptance of the content of a communication (8, 10).

A difficulty in generalizing too confidently from such findings arises when one considers that active rehearsal of this nature is assumed to reinforce the content of the communication and, hence, might be expected to facilitate attitude change only in those situations where the participants concur with the communication content. In cases where the communication tends to counteract existing attitudes in the audience, it is questionable that active participation would provide the kind of conditions that would enhance the effects of the communication. It seems particularly questionable with respect to groups that are unified in their antagonism to the communication content. Prediction of the direction, extent, and stability of attitude change in groups predisposed in different ways toward a controversial film thus poses a problem. What, for example, would be the nature of attitude changes in groups composed homogeneously of individuals who have a positive, negative, or neutral bias toward the content of the communication when opportunity for discussion of the material is extended or withheld? The experimental literature in the field of communication and persuasion provides no basis for making clear-cut predictions under several conditions described.

In one sense, as Hovland, Janis, and Kelly (6) have pointed out, the effectiveness of a persuasive communication is a matter of learning. Any attitude change following exposure to a film, for example, may reflect the amount of new information conveyed by the film to the audience. Such change seems to be most effective with respect to those items that are covered specifically in the film content (7). A relationship between attitude and retention of information has been found by Doob (4), who reports that intensity rather than direction of attitude facilitates retention of the content of a communication. If such is the case, then the familiar principles of learning do not apply without emendation to persuasive communication, since superior retention for pleasant, or acceptable, as opposed to unpleasant, or unacceptable, material has been reported in several studies (5, 17, 18). A further problem intrudes here, namely the effects of initial attitude upon learning among the recipients of a persuasive communication. Does initial attitude influence the amount of learning from a relevant communication, and is attitude change related to such learning?

The problems suggested here become even more complex when one adds the variable of active participation to the audience settings, since it seems well-established that participation in a communication situation enhances learning on the part of those who participate (9, 13). The present study was designed to measure the relative influence of the factors of initial attitude and learning upon participation, the retention of information, and the direction, extent, and stability of attitude change.

---

[1] This research was supported in part by a special grant from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service. The assistance of Dr. Willard Vaughan and Mr. Clagett Smith, our co-researchers, is gratefully acknowledged. Dr. Ray C. Hackman assisted in the design and statistical analysis. This paper is based upon portions of the doctoral dissertation of the first author.

[2] Now with the Human Resources Research Office, Washington, D. C.

METHOD

Subjects

Approximately 400 students from two high schools[3] responded to a slightly modified version of the California Ethnocentrism (E) Scale. In each school, 27 students were selected from each extreme and from the middle of the distribution of E scores, which closely approximated a Gaussian curve. The three levels of ethnocentrism represented will be referred to as low E, high E, and middle E, indicating the portion of the score distribution from which the *S*s were chosen. This preliminary testing took place about one month prior to the experimental sessions. Means and variances were determined for the students of the two schools and no significant differences between schools were found for either of these measures.

Procedure

Within each category of ethnocentrism, nine *S*s were assigned at random to each of three experimental conditions, designated as film-discussion, film-alone, and control. This was done within schools, so that a total of 18 groups, or 162 *S*s, representing three degrees of ethnocentrism participated in the experiment. The film-discussion groups first viewed the film "The High Wall," which treats group prejudice as a communicable disease and traces its origins in the family and community. The members of these groups, representing low, high, and intermediate degrees of ethnocentric disposition, were given thirty minutes in which to discuss the film and related issues. The group members were free to participate in the discussion or not, and the discussion leader assumed a permissive, or nondirective, role. The same discussion leader served for all of the groups. The discussions were terminated either at the end of half an hour or when it appeared that the group members had no more comments to offer. At the conclusion of the discussion, the *S*s again answered the E scale.

Members of the film-alone groups viewed the film under comparable conditions but did not engage in discussion of it. They again filled out the E scale upon conclusion of the film showing. The control groups responded to the E scale at the same times as the experimental groups but without any intervening exposure to either the film or a discussion.

In order to determine the amount of learning of film content by the *S*s, a 33-item multiple-choice Information test covering factual material in the film was devised through pretesting and item-analysis. The items on the test covered such points as identification of characters in the film and details of the action in certain scenes. The items did not deal directly with the problem of prejudice. The Information test, therefore, was not a measure of attitude-related material but, rather, was an index of the sheer amount of detail retained from exposure to the film. Pretesting with this form would have been meaningless, since the pattern of responses would necessarily be random for *S*s who had not viewed the film. This test was administered to each of the experimental groups (film-discussion and film-alone) immediately after the second administration of the E scale, following presentation of the film. Both the E scale and the Information test were readministered a final time one month after the film showing in order to determine retention of information and stability of attitude change during this interval. The control groups, of course, did not take the Information test at any time, since it dealt exclusively with the content of the film.

All of the discussions were tape-recorded, and the recording equipment was present whether a discussion was to follow the film or not. None of the apparatus was concealed from the discussion groups, but they were assured that all records would be held in confidence. Two other individuals were present in the experimental situation; they were the "spotters" who, from the rear of the room, recorded participation of an *S* in terms of *S*'s seat by row and seat number. When an individual participated, his location was noted by both spotters in terms of these references. Thus, it was possible to maintain a running account of participation by each *S*, and to identify each comment with its source upon typed, verbatim transcripts of the recordings. This procedure is described in detail elsewhere (11).

RESULTS

Attitude Change

The measure used to assess attitude change was a difference score determined by subtracting the post treatment score from the initial score on the E scale. Since it is possible that a change of position on an attitude scale may be related to one's initial attitude, the correlation coefficient between initial attitudes and difference scores was computed and found to be 0.16, indicating that less than three per cent of the variability in the difference scores could be accounted for by the initial scores.

---

[3] Miss Rowenetta Allen, Supervisor of Instruction of Prince George's County, Maryland, kindly granted permission for inviting two high schools to participate in this research. Mr. Allan Chotiner and Mr. Fred Wampler of High Point Senior High School, and Mr. Kalman Vozar and Mr. Ralph Angel of Bladensburg Senior High School cooperated generously in all phases of the study.

TABLE 1

ANALYSIS OF VARIANCE OF ATTITUDE CHANGES

Source	df	Sums of Squares	Mean Squares	F^a
Predisposition (P)	2	779.4	389.7	3.62
Treatments (T)	2	6506.9	3253.5	30.18**
P × T	4	431.0	107.8	5.36**
Within groups	153	3068.1	20.1	
Total	161	10785.4		

[a] A single asterisk indicates rejection of the null hypothesis at the .05 level of confidence, and the double asterisk indicates rejection at the .01 level.

TABLE 2

MEAN ATTITUDE CHANGE SCORES[a]

	Low E	Middle E	High E
Control	32.5	35.9	37.8
Film-alone	44.9	48.5	54.7
Film-discussion	47.5	49.5	47.6

[a] For ease in computation a constant of +40 was added to all difference scores.

Data from the two high schools were combined for all statistical analyses. After Bartlett's test provided no basis for rejecting the assumption of homogenous variances, an analysis of variance was performed on the attitude change scores. As shown in Table 1, the F values for the treatment × predisposition interaction and treatment as a main effect were significant at the .01 level of confidence. Table 2 shows the mean values for each treatment and predisposition. Among the high E individuals differences in attitude change between the film-alone and film-discussion conditions were significant at the .05 level of confidence. The differences between the film-alone and film-discussions treatment for the middle E and low E groups were not significant. For the film-alone condition the high E groups showed significantly more change than the other two groups, but for the film-discussion treatment there were no significant differences among predisposition groups. It is evident, then, that giving the high E groups the opportunity to discuss the film counteracted some of the effects of the film, while for the low E groups a reversal of this trend was noted. The significance of treatment as a main effect was due to differences between control groups and both experimental groups; there was no significant difference between the combined film-discussion groups and the film-alone groups.

Fig. 1 presents the mean values for the initial, posttreatment and delayed administration of the E scale for all treatments. Since predisposition was not associated with attitude change, the various predisposition groups are combined according to treatment. The attitude change scores (initial-minus-post) for the film-discussion groups and for the film-alone groups both differed significantly from the control groups at the .01 level of confidence, but the experimental groups did not differ with respect to attitude change. In short, significant reductions in personal ethnocentrism were brought about by both the film-discussion and film-alone treatments, while the control groups showed no change in ethnocentric attitudes.

*S*s were retested with the E scale one month after treatment, and an analysis of variance was performed on the differences between the initial scores and the delayed scores on the E scale. Inasmuch as some of the *S*s were absent at the time of the delayed E scale, their scores were estimated by a procedure suggested by Cochran and Cox (3). Before performing this analysis, the correlation coefficient between this difference score and the initial score was determined and found to be 0.04, indicating that only 0.16% of the variability in the difference scores could be attributed to the initial

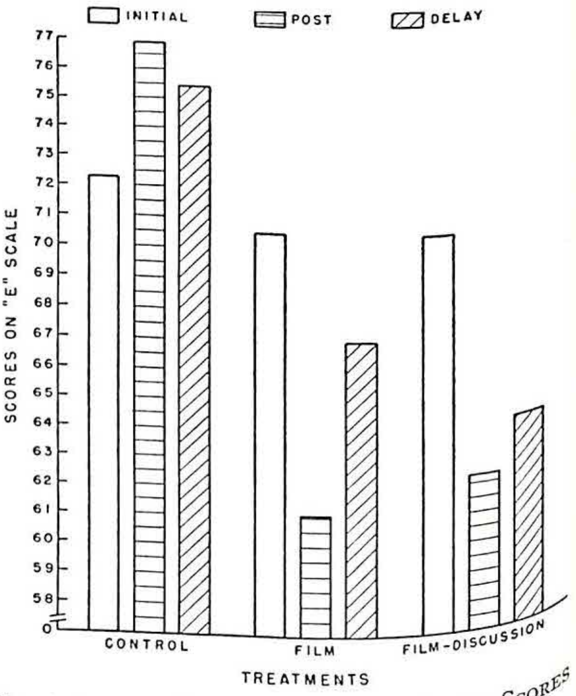


FIG. 1. INITIAL, POSTTREATMENT, AND DELAY SCORES ON THE "E" SCALE FOR EACH TREATMENT

TABLE 3
ANALYSIS OF VARIANCE OF STABILITY OF ATTITUDE CHANGE

Source	*df*	Sums of Squares	Mean Squares	*F*
Predisposition (P)	2	40.1	20.0	—
Treatments (T)	2	2563.3	1281.6	9.52**
P × T	4	625.3	156.3	1.16
Within groups	139	18709.7	134.6	
Total	147	21938.4		

scores. An examination of this analysis, shown in Table 3, shows that the obtained *F* value for treatment was significant at the .01 level of confidence. Fig. 1 presents the delay scores for each treatment. Differences between the retention of attitude change for the film-alone and film-discussion groups differed significantly from the controls at the .01 level of confidence, but these two groups did not differ from each other when compared with respect to initial score minus delayed score.

A further comparison of the scores represented in Fig. 1 was made. Within each experimental condition, the scores on the posttreatment and delayed administration of the E scale were compared. Neither the control groups nor the film-discussion groups differed significantly on these two measures, indicating that for these two conditions there was no significant change one month after treatment from posttreatment scores. However, the film-alone groups showed a significant change on these two measures. At the time of the delayed measurement, the film-alone groups had moved significantly from their posttreatment score position in the direction of their pre-experimental position.

It was previously indicated that on the posttreatment measure the film-alone groups and the film-discussion groups did not differ significantly. A breakdown of the film-discussion groups was made into active and passive participants, with the active participants defined as those individuals who made one or more comments in the discussion regardless of length, and the passive participants defined as those individuals who made no contribution to the discussion. Of the 54 *S*s who comprised the film-discussion groups, 37 were active participants and 17 were passive participants. Fig. 2 shows the mean values for three degrees of participation on each administration of the E scale. None of the changes from initial to posttreatment scores differed significantly from one another.

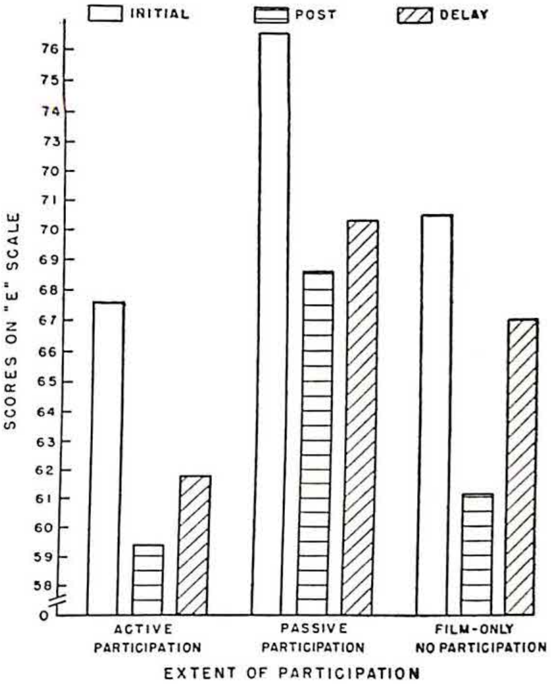


FIG. 2. INITIAL, POSTTREATMENT, AND DELAY SCORES ON THE "E" SCALE FOR EACH LEVEL OF PARTICIPATION

A comparison of the posttreatment and delayed E-scale positions had shown that the film-alone groups moved significantly from their posttreatment scores in the direction of their original position, while the control and film-discussion groups showed no such change. Is this difference attributable to greater attitude change among the active participants in the film-discussion groups, or does it result from the process of discussion influencing all members present, whether active or passive? In order to provide an answer to this question, the active and passive participants were compared on their posttreatment and delayed scores. Fig. 2 shows these differences graphically. The differences between the posttreatment and delayed E scores were not significant for either active or passive participants. That is, neither the active nor the passive discussion participants changed significantly from their posttreatment measures after an interval of one month. The film-alone groups on the other hand, regressed toward their earlier position. The stability of attitude change characterizing both the active and passive participants suggests that presence during a discussion is as effective as active participation in generating greater permanency of attitude change.

Learning

The relationship of amount of information learned from the film to the major independent variables of the experiment was also evaluated by analysis of variance. Since the relationship between learning and participation, however, was ascertainable only in those groups that held discussions, that analysis is presented separately.

Examination of Table 4 shows that the *F* value for predisposition was significant at the .01 level of confidence. The mean scores for the three predisposition groups were: low E, 24.8; middle E, 22.3; high E, 19.8. All differences between these means were significant at the .01 level of confidence, indicating an inverse relationship between initial degree of ethnocentrism and learning of information specific to the film. Since treatments were not a significant source of variability, it is apparent that discussion following the film communication did not facilitate recall of details from the film.

TABLE 4

ANALYSIS OF VARIANCE OF SCORES ON THE INFORMATION TEST

Source	*df*	Sums of Squares	Mean Squares	*F*
Predisposition (P)	2	450.7	225.4	15.65**
Treatment (T)	1	16.3	16.3	1.13
P × T	2	18.7	9.4	
Within groups	102	1468.0	14.4	
Total	107	1953.7		

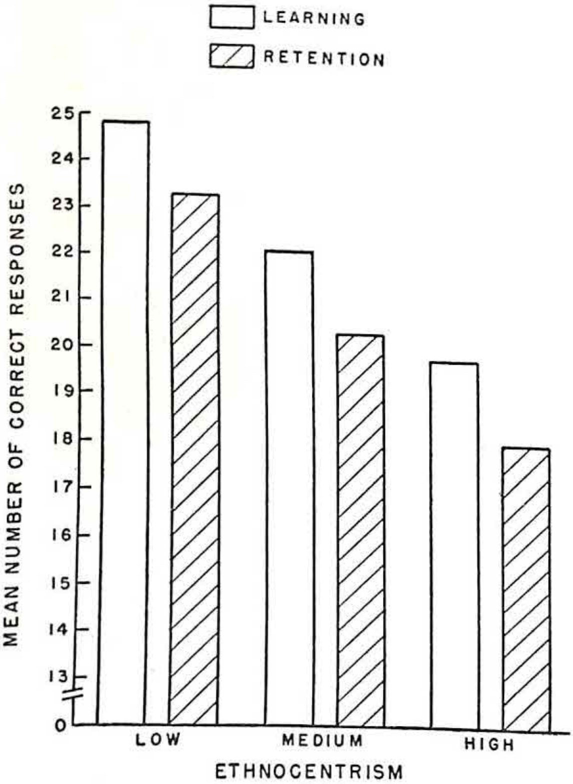


FIG. 3. MEAN SCORES ON THE LEARNING AND RETENTION TEST FOR EACH PREDISPOSITION GROUP

Fig. 3 presents the mean learning and retention scores of each predisposition group. It is evident that individuals who scored high on the E scale learned the least from the film, while those who were low on the E scale learned the most from the film. The middle E groups fell in between these extremes. Those individuals, in short, who were assumed on the basis of the E scores to be favorably disposed toward the content of the communication learned significantly more factual material from it than did those *S*s who were antagonistic to the film's theme, even though the information learned was not directly related to the topic of prejudice. That this relationship was not confounded with differences in intelligence among the *S*s was attested by the fact that the correlation between E scores and scores on the California Mental Maturity Test was not significant.

An analysis of variance was also performed on the retention of information scores to determine whether the experimental variables were reflected in differential retention of film content among the *S*s one month after the first testing. The scores of the absentees were again estimated by the procedure suggested by Cochran and Cox (3). Results of the analysis of variance which are given in Table 5 show that predisposition is significant at the .01 level. The differences in retention scores among the three predisposition conditions are presented graphically in Fig. 3. An increased degree of ethnocentrism, in general, is associated with a greater decrement in retention of information. It should be noted, however, that a correlation of .73 ($P < .01$) was found between learning and retention, so that the differences in retention of information may be ascribed to amount of material initially learned rather than to predisposition as such.

Finally, an effort was made to determine

TABLE 5
ANALYSIS OF VARIANCE OF INFORMATION RETENTION SCORE

Source	*df*	Sum of Squares	Mean Squares	*F*
Predisposition (P)	2	483.6	241.8	15.80**
Treatment (T)	1	35.6	35.6	—
P × T	2	90.7	45.4	2.97
Within groups	94	1436.5	15.3	
Total	99	2046.4		

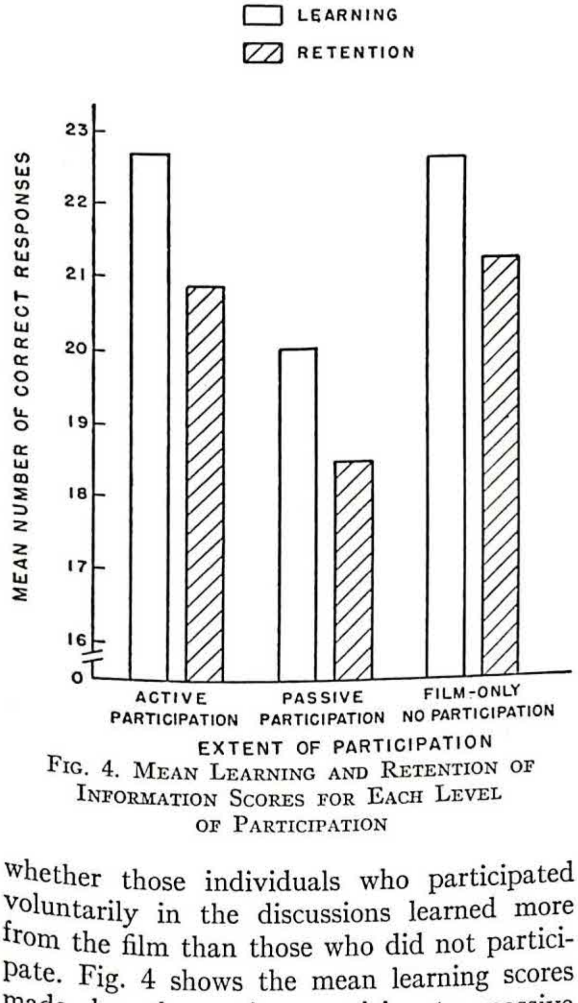


FIG. 4. MEAN LEARNING AND RETENTION OF INFORMATION SCORES FOR EACH LEVEL OF PARTICIPATION

whether those individuals who participated voluntarily in the discussions learned more from the film than those who did not participate. Fig. 4 shows the mean learning scores made by the active participants, passive participants, and film-alone *S*s. Evaluation by means of *t* tests revealed no significant differences between the active participants and members of the film-alone groups. Both the active participants and the film-alone *S*s, however, learned significantly more from the film than the passive discussion participants. This advantage persisted on the retention measure one month later.

In an attempt to determine whether attitude change was related to the amount of information learned, a correlation was computed between these two variables. The obtained value of .03 indicated that there was no significant relationship between learning of film details and attitude change.

DISCUSSION

Attitude change in response to a persuasive sound film was a joint function of opportunity for discussion of the film and initial attitude. Predisposition alone was not related to attitude change. It had been expected that in the low E groups a greater effect might be obtained with those members who held discussions, since these individuals had views that were essentially in accord with those contained in the film. (Even the low Es could show attitude change since the mean score of these *S*s was 26.7 points below the ceiling of the E scale.) The discussion might reasonably have been expected to reinforce the initial attitudes of these *S*s and thus facilitate a greater change in attitude than that produced among the low Es who did not hold discussions. In the high E groups, on the other hand, a possibility existed that expression of opinions at variance with those presented in the film might mitigate the film's effect, and that the nondiscussion groups might therefore show greater attitude change. Among the neutral or middle E *S*s, no difference was expected between the film and film discussion conditions.

These hypotheses were generally supported by the data. Although the difference in attitude change in the low E groups under film-alone and film-discussion conditions was not significant, the trend was in the expected direction. The high Es showed significantly less change under discussion conditions. In general, it may be suggested that discussion of a persuasive communication does facilitate attitude change among those initially favorable to the communication content. On the other hand, discussion of a communication by homogeneous groups that are antagonistic to the content detracts from the effectiveness of the communication. This result may be attributed to the fact that discussion provides an opportunity for such individuals to express their initial views and thus partially to negate the persuasive effects that have been generated. Examination of the discussion transcripts supports this interpretation. The high Es devoted most of the discussion periods to expression of their antipathies toward Negroes, while the low E and middle E groups tended to examine the general problems of group prejudice raised by the film. The middle *S*s having no clear-cut opinion either pro or con held very apathetic discussions and, therefore, neither reinforced nor mitigated the effects of the film.

While there were no differences in attitude change between the film-alone and film-discussion groups immediately following the communication, the film-alone groups regressed significantly from their posttreatment positions when measured one month later; the film-discussion groups and the controls showed no such change. The discussions probably gave those *S*s present an opportunity to learn how other members of their group perceived the film and permitted the convergence of views toward a common norm. Apparently influenced by such a consensus effect, the film-discussion groups maintained their posttreatment attitude change. The film-alone groups, not having experienced this interaction, regressed towards their initial scores on the delayed measurement. Thus, participation in the form of group discussion following a communication may have some value in maintaining attitude changes produced by the communication, even when the initial gains are not measurably affected. Our results indicated that extent of participation was not a factor in determining initial attitude change. Participation, however, did influence the permanence of posttreatment attitude changes. On the delayed measurement neither the active nor the passive participants regressed significantly from their posttreatment scores, while the film-alone groups did show such a regression. Stability of the attitude change, therefore, cannot be attributable to active participation alone, but must be due in part to mere presence in a group discussion. Although active participation has been defined as verbal commitment, the possibility of nonverbal communication during the discussion must be considered. Reinforcement of opinion change may be the result of a number of aspects of the group discussion situation not limited to verbal interchange.

A decreasing amount of factual material was learned from the film in moving from the groups low in ethnocentrism to those high in ethnocentrism. This finding implies that those *S*s whose attitudes were assumed to be congruent with the content of the film learned the most from it, while the *S*s whose attitudes were in conflict with the film content learned the least amount of factual material. Our findings thus substantiate the hypothesis that learning from a persuasive communication is in part a function of initial attitude toward the communication content.

A content analysis of the discussion materials indicated that discussion in the low E groups tended to center about the arguments presented by the film, whereas discussion in the high E groups departed almost completely from the film. While it was not determined how much information in the film was reviewed in each discussion, the fact that more comments in the low E groups were related to the film increased the probability that specific informational items would be mentioned. Actually, the effects of discussion content and initial attitude are confounded in the discussion groups. In the film-alone groups, however, no such confounding of these two variables is present, and here the low Es also learned significantly more from the film than the high Es. In addition, the analysis of variance of Information test scores showed that treatment (discussion vs. no discussion) was not a significant source of variability. These considerations indicate that discussion of a communication does not necessarily alter the amount of information acquired from the communication. In the present instance, the discussion comments were not such as to involve rehearsal of factual content of the film. The probability that the amount of practice furnished by the discussion was slight would account for the lack of differences in learning between the film-discussion and film-alone groups.

Although there were no significant differences on the initial administration of the E scale between the active and passive participants, a significant difference was observed between these groups in amount of information learned. The fact that the active participants learned more than the passive participants agrees with the findings of Kurtz and Hovland (9). In light of this, one might predict that the film-alone *S*s, having no opportunity for participation, would learn less from the film than either the active or passive members of the discussion groups. However, the results showed that the film-alone *S*s learned significantly more than the passive participants, and, in fact, did not differ significantly from the active participants. These findings seem to indicate that participation depended on amount of material learned, since the film-alone groups undoubtedly contained some individuals who would have participated actively in discussion given the opportunity. The presence of these potential discussants would tend to raise the average learning score in the film-alone groups above the mean of the passive members of the discussion groups. In other words, the mean learning score for the film-alone groups represents the joint contribution of potentially active and passive discussion participants.

Data reported by McGinnies and Vaughan (12) support the contention that participation in a free discussion depends upon the amount of relevant information possessed by *S*s. Their results indicated that the *S*s who participated in discussions following mental health films rated themselves higher on a self-rating scale of mental health information than those who did not participate in the discussion. They also reported that the discussants had seen significantly more mental health films than the nondiscussants. Participation in a discussion appears to depend in part upon knowledge of the problem.

At the time of the delayed administration of the Information test, the low E groups obtained the highest scores, and the high E groups achieved the poorest retention scores. Although predisposition, or initial attitude, was not a significant source of variance when the retention scores were adjusted for initial learning, the predisposition × treatment interaction was nearly significant. In other words, the observed differences in retention of information among the several predisposition groups probably depended upon the occurrence or nonoccurrence of a discussion following the film. For the low E and high E groups, the more effective treatment was the film-discussion situation.

Retention of information was obviously unrelated to extent of participation in discussion, since the film-alone groups achieved retention scores equal to those of the active discussion participants and higher than those of the passive participants. Furthermore, no significant differences existed between the active and passive participants in this respect.

SUMMARY AND CONCLUSIONS

This study investigated the effects of ethnocentric attitudes and discussion participation upon the degree and stability of learning and attitude change in response to a persuasive communication. Twelve experimental groups, each composed of nine high school students, viewed a sound film "The High Wall." Six of the groups held discussions of the film, while the remaining six did not. The groups were formed on the basis of scores on the California Ethnocentrism Scale, so that two groups under each of the experimental conditions represented low, middle, and high scores on the E scale. Following experimental treatment, film-discussion or film-alone, all of the group members were tested for attitude change and amount of factual information learned from the film. Six control groups were given pre- and posttests on the questionnaires without intervening exposure to either the film or discussion. One month following the experimental treatments, all groups were retested on the E scale, and the experimental groups were also readministered the Information test.

The major findings of this experiment may be summarized as follows:

1. Significant reductions in ethnocentrism, as measured by the E scale, were produced in both the film-discussion and film-alone groups. For prejudiced *S*s, the effect was significantly less in the discussion groups than in the nondiscussion groups. Ethnocentric predisposition alone, however, did not affect the extent of attitude change induced by the film.
2. Stability of attitude change was a function of experimental treatment. The active and passive members of discussion groups, when tested one month later, had largely retained their posttreatment attitude gains, while the film-alone groups had regressed significantly toward their original attitudes.
3. Amount of information learned from the film was related to initial attitude. Those individuals who were low in ethnocentrism learned more from the film than those high in ethnocentrism. This trend persisted in the retention measures taken a month later.
4. Active discussion participants learned more from the film than passive participants.

Analysis of this finding led to the conclusion that extent of participation in discussion depends upon the amount of learning that has taken place.

REFERENCES

1. BAVELAS, A. Role playing and management training. *Sociatry*, 1947, **1,** 183–191.
2. COCH, L., & FRENCH, R. P., JR. Overcoming resistances to change. *Hum. Relat.*, 1948, **1,** 512–532.
3. COCHRAN, W. G., & COX, G. M. *Experimental design.* New York: Wiley, 1950.
4. DOOB, L. W. Effects of initial serial positions and attitude upon recall conditions of low motivation. *J. abnorm. soc. Psychol.*, 1953, **48,** 199–205.
5. EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36,** 34–50.
6. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.
7. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication.* Princeton: Princeton Univer. Press, 1949.
8. JANIS, I. L., & KING, B. T. The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49,** 211–218.
9. KURTZ, K. H., & HOVLAND, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, **45,** 157–164.
10. LEWIN, K. Group decision and social change. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Henry Holt, 1947. Pp. 330–344.
11. McGINNIES, E. A method for matching anonymous questionnaire data with group discussion material. *J. abnorm. soc. Psychol.*, 1956, **52,** 139–140.
12. McGINNIES, E., & VAUGHAN, W. Some biographical determiners of participation in group discussion. *J. appl. Psychol.*, 1957, **41,** 179–185.
13. MICHAEL, D. M., & MACCOBY, N. Factors influencing verbal learning from films under varying conditions of audience participation. *J. exp. Psychol.*, 1953, **46,** 411–418.
14. PETERSON, R. C., & THURSTONE, L. L. *Motion pictures and the social attitudes of children.* New York: Macmillan, 1933.
15. RAMSEYER, L. L. A study of the influence of documentary films on social attitudes. Unpublished doctoral dissertation, Ohio State Univer., 1938.
16. RATHS, L. E., & TRAGER, F. N. Public opinion and crossfire. *J. educ. Soc.*, 1948, **21,** 345–368.
17. SHARP, A. A. An experimental test of Freud's doctrine of the relation of hedonic tone to memory revival. *J. exp. Psychol.*, 1938, **22,** 395–418.
18. WATSON, W. S., & HARTMAN, G. W. The rigidity of a basic attitudinal frame. *J. abnorm. soc. Psychol.*, 1939, **34,** 314–335.

Received April 15, 1957.

SOME FACTORS IN SELF-DISCLOSURE[1]

SIDNEY M. JOURARD[2] AND PAUL LASAKOW

University of Alabama Medical Center and University of Alabama Birmingham Center

THE present paper describes a questionnaire method for measuring the amount and content of self-disclosure to selected "target-persons," and reports the results of three exploratory studies. Self-disclosure refers to the process of making the self known to other persons; "target-persons" are persons to whom information about the self is communicated.

The process of self-disclosure has been studied by others from various points of view. Block (1) and Block and Bennett (2) have demonstrated that the content of communication about the self is a function of variations in "own-role." Lewin (7) noted differences between typical Germans and Americans regarding their readiness to confide personal information to others. Jourard (6) has suggested that accurate portrayal of the self to others is an identifying criterion of healthy personality, while neurosis is related to inability to know one's "real self" and to make it known to others. Characterological studies of Fromm (3), Riesman (9), and Horney (4) have called attention to a tendency common among persons in our society, to misrepresent the self to others. This tendency is central to the "marketing personality," the "other-directed character," and the "self-alienated" individual, as these have been described by their respective authors. Since much of social science is founded upon the self-disclosures of respondents, the conditions and dimensions of self-disclosure bear directly upon the validity of many purported facts in the social sciences.

From the foregoing, it may be concluded that systematic analysis of self-disclosure holds promise of yielding information that is relevant to diverse areas of theory and method.

The following questions were proposed for investigation:

1. Do subjects (*S*s) vary in the extent to which they disclose themselves to different target-persons, for example, mother, father, male friend, and female friend? What is the effect of the *S*s' marital status on self-disclosure to parents and friends? What is the effect of the *S*s' feelings and attitudes toward particular target-persons upon self-disclosure to them? The last question was investigated only with respect to the relationship between *S*s' disclosure of self to parents, and their feelings and attitudes toward their parents.

2. Are there differences between categories of information about the self (*aspects* of self) with respect to self-disclosure? Do *S*s tend to disclose some aspects of self more fully than others?

3. Are there differences ascertainable between Negro and white *S*s with respect to self-disclosure?

4. Are there sex differences regarding self-disclosure?

METHOD

Instruments

The Self-disclosure Questionnaire. A sixty-item questionnaire was devised. As can be seen in Table 1, the items are classified in groups of ten within each of six more general categories of information about the self (*aspects*). *S*s were given the following instructions for completing the questionnaire:

> The answer-sheet which you have been given has columns with the headings "Mother," "Father," "Male Friend," "Female Friend," and "Spouse." You are to read each item on the questionnaire, and then indicate on the answer-sheet the extent that you have talked about that item to each person; that is, the extent to which you have made yourself known to that person. Use the rating-scale that you see on the answers sheet to describe the extent that you have talked about each item.

The self-disclosure rating-scale was as follows:

> 0: Have told the other person nothing about this aspect of me.
>
> 1: Have talked in general terms about this item. The other person has only a general idea about this aspect of me.
>
> 2: Have talked in full and complete detail about this item to the other person. He knows me fully in this respect, and could describe me accurately.
>
> X: Have lied or misrepresented myself to the other person so that he has a false picture of me.

The numerical entries were summed (*X*'s were counted

---

[1] The authors are indebted to Drs. A. J. Riopelle, Emory University, and O. Lacey, University of Alabama, for statistical advice.

[2] Now at the University of Florida, Gainesville.

TABLE 1

THE SELF-DISCLOSURE QUESTIONNAIRE

Attitudes and opinions
1. What I think and feel about religion; my personal religious views.
2. My personal opinions and feelings about other religious groups than my own, e.g., Protestants, Catholics, Jews, atheists.
3. My views on communism.
4. My views on the present government—the president, government policies, etc.
5. My views on the question of racial integration in schools, transportation, etc.
6. My personal views on drinking.
7. My personal views on sexual morality—how I feel that I and others ought to behave in sexual matters.
8. My personal standards of beauty and attractiveness in women—what I consider to be attractive in a woman.
9. The things that I regard as desirable for a man to be—what I look for in a man.
10. My feelings about how parents ought to deal with children.

Tastes and interests
1. My favorite foods, the ways I like food prepared, and my food dislikes.
2. My favorite beverages, and the ones I don't like.
3. My likes and dislikes in music.
4. My favorite reading matter.
5. The kinds of movies that I like to see best; the TV shows that are my favorites.
6. My tastes in clothing.
7. The style of house, and the kinds of furnishings that I like best.
8. The kind of party, or social gathering that I like best, and the kind that would bore me, or that I wouldn't enjoy.
9. My favorite ways of spending spare time, e.g., hunting, reading, cards, sports events, parties, dancing, etc.
10. What I would appreciate most for a present.

Work (or studies)
1. What I find to be the worst pressures and strains in my work.
2. What I find to be the most boring and unenjoyable aspects of my work.
3. What I enjoy most, and get the most satisfaction from in my present work.
4. What I feel are *my* shortcomings and handicaps that prevent me from working as I'd like to, or that prevent me from getting further ahead in my work.
5. What I feel are my special strong points and qualifications for my work.
6. How I feel that my work is appreciated by others (e.g., boss, fellow-workers, teacher, husband, etc.)
7. My ambitions and goals in my work.
8. My feelings about the salary or rewards that I get for my work.
9. How I feel about the choice of career that I have made—whether or not I'm satisfied with it.
10. How I really feel about the people that I work for, or work with.

Money
1. How much money I make at my work, or get as an allowance.
2. Whether or not I owe money; if so, *how much*.

TABLE 1—*Continued*

3. Whom I owe money to at present; or whom I have borrowed from in the past.
4. Whether or not I have savings, and the amount.
5. Whether or not others owe me money; the amount, and who owes it to me.
6. Whether or not I gamble; if so, the way I gamble, and the extent of it.
7. All of my present sources of income—wages, fees, allowance, dividends, etc.
8. My total financial worth, including property, savings, bonds, insurance, etc.
9. My most pressing need for money right now, e.g., outstanding bills, some major purchase that is desired or needed.
10. How I budget my money—the proportion that goes to necessities, luxuries, etc.

Personality
1. The aspects of my personality that I dislike, worry about, that I regard as a handicap to me.
2. What feelings, if any, that I have trouble expressing or controlling.
3. The facts of my present sex life—including knowledge of how I get sexual gratification; any problems that I might have; with whom I have relations, if anybody.
4. Whether or not I feel that I am attractive to the opposite sex; my problems, if any, about getting favorable attention from the opposite sex.
5. Things in the past or present that I feel ashamed and guilty about.
6. The kinds of things that just make me furious.
7. What it takes to get me feeling real depressed and blue.
8. What it takes to get me real worried, anxious, and afraid.
9. What it takes to hurt my feelings deeply.
10. The kinds of things that make me especially proud of myself, elated, full of self-esteem or self-respect.

Body
1. My feelings about the appearance of my face—things I don't like, and things that I might like about my face and head—nose, eyes, hair, teeth, etc.
2. How I wish I looked: my ideals for overall appearance.
3. My feelings about different parts of my body—legs, hips, waist, weight, chest, or bust, etc.
4. Any problems and worries that I had with my appearance in the past.
5. Whether or not I now have any health problems—e.g., trouble with sleep, digestion, female complaints, heart condition, allergies, headaches, piles, etc.
6. Whether or not I have any long-range worries or concerns about my health, e.g., cancer, ulcers, heart trouble.
7. My past record of illness and treatment.
8. Whether or not I now make special efforts to keep fit, healthy, and attractive, e.g., calisthenics, diet.
9. My present physical measurements, e.g., height, weight, waist, etc.
10. My feelings about my adequacy in sexual behavior—whether or not I feel able to perform adequately in sex-relationships.

as zeros), yielding totals which constituted the self-disclosure scores.

Seventy white unmarried college students of both sexes were tested for self-disclosure to Mother, Father, Male Friend, and Female Friend, in a study of reliability. Since the questionnaire included 60 items, and there were four target-persons, a total of 240 entries were made by each *S*. These 240 entries were divided into halves by the odd-even method, and the subtotal sums were correlated with each other. The resultant *r*, corrected, was .94, indicating that the *S*s were responding consistently to the questionnaire over all target persons, and all aspects of self.

Parent-Cathexis Questionnaires. Mother-cathexis and Father-cathexis questionnaires, fully described elsewhere (5), were employed to test *S*s' feelings toward their parents. The *S*s rated their feelings about 40 parental traits, e.g., *sense of humor*, *temper*, *ability to make decisions*, in accordance with the following scale:

1: Have strong positive feelings; like very much.
2. Have moderate positive feelings.
3. Have no feelings one way or the other.
4. Have moderate negative feelings.
5. Have strong negative feelings; dislike very much.

High scores indicated negative feelings toward the parents, while low scores signified positive feelings.

Subjects

The *S*s included in the studies to be reported were taken from larger samples drawn from three Alabama college populations: two white liberal arts colleges, a Negro liberal arts college, and a school of nursing located at a medical school. For the combined sample 300 white and Negro liberal arts college sophomores and juniors were obtained, and 55 white nursing students. All *S*s were tested in groups by an examiner of the same race.

For the purpose of analysis, the following randomly selected subsamples were drawn from the combined samples:

1. From the 300 liberal arts students, a subsample of 10 white males, 10 white females, 10 Negro males, and 10 Negro females was drawn for the study of differences in self-disclosure associated with race, sex, targets, and aspects of self. All *S*s were unmarried, and in all cases, the parents were living. Mean ages were: white males, 21.70, *SD* 2.00; white females, 20.30, *SD* .90; Negro males, 22.10, *SD* 2.02; and Negro females, 20.40, *SD* .45. All *S*s had been tested for self-disclosure to Mother, Father, Male Friend, and Female Friend.

2. From all the white respondents in the combined sample, a subsample of 10 married male and 10 married female *S*s was drawn for comparison with the first subsample of 10 unmarried males and 10 unmarried females to test the effects of marriage on self-disclosure patterns. These *S*s had indicated self-disclosure to Mother, Father, Same-sex Friend, and Spouse. Mean age for the married males was 23.40, *SD* 1.43, and for the married females, 20.60, *SD* 2.42.

3. Thirty-one unmarried nursing students comprised the third sample, used to examine the relationship between parent-cathexis and self-disclosure to the parents. Mean age for this group was 18.59, *SD* 3.53.

The data were analyzed according to Lindquist's

TABLE 2

ANALYSIS OF VARIANCE OF SELF-DISCLOSURE SCORES OF WHITE AND NEGRO, MALE AND FEMALE SUBJECTS

Source	Mean Square	*df*	Error Term	*F*
Between groups	892.33	3	Error (b)	6.78***
Males vs. Females	573.50	1		4.36*
Whites vs. Negroes	2076.82	1		15.78***
Sex × Race	126.67	1		
Error (b)	131.61	36		
Total *df* Between		39		
(Within groups comparisons)				
Between Target-Persons	557.06	3	Error$_1$ (w)	10.89***
Between Aspects of Self	952.01	5	Error$_2$ (w)	40.39***
Target × Aspect	87.49	15	Error$_3$ (w)	15.38***
Group × Target	128.11	9	Error$_1$ (w)	2.50*
Group × Aspect	32.16	15	Error$_2$ (w)	1.36
Group × Target × Aspect	9.61	45	Error$_3$ (w)	1.69**
Total *df* Within		92		
Error$_1$ (w)	51.17	108		
Error$_2$ (w)	23.57	180		
Error$_3$ (w)	5.69	540		
Total Error *df*		828		
Total *df*		959		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

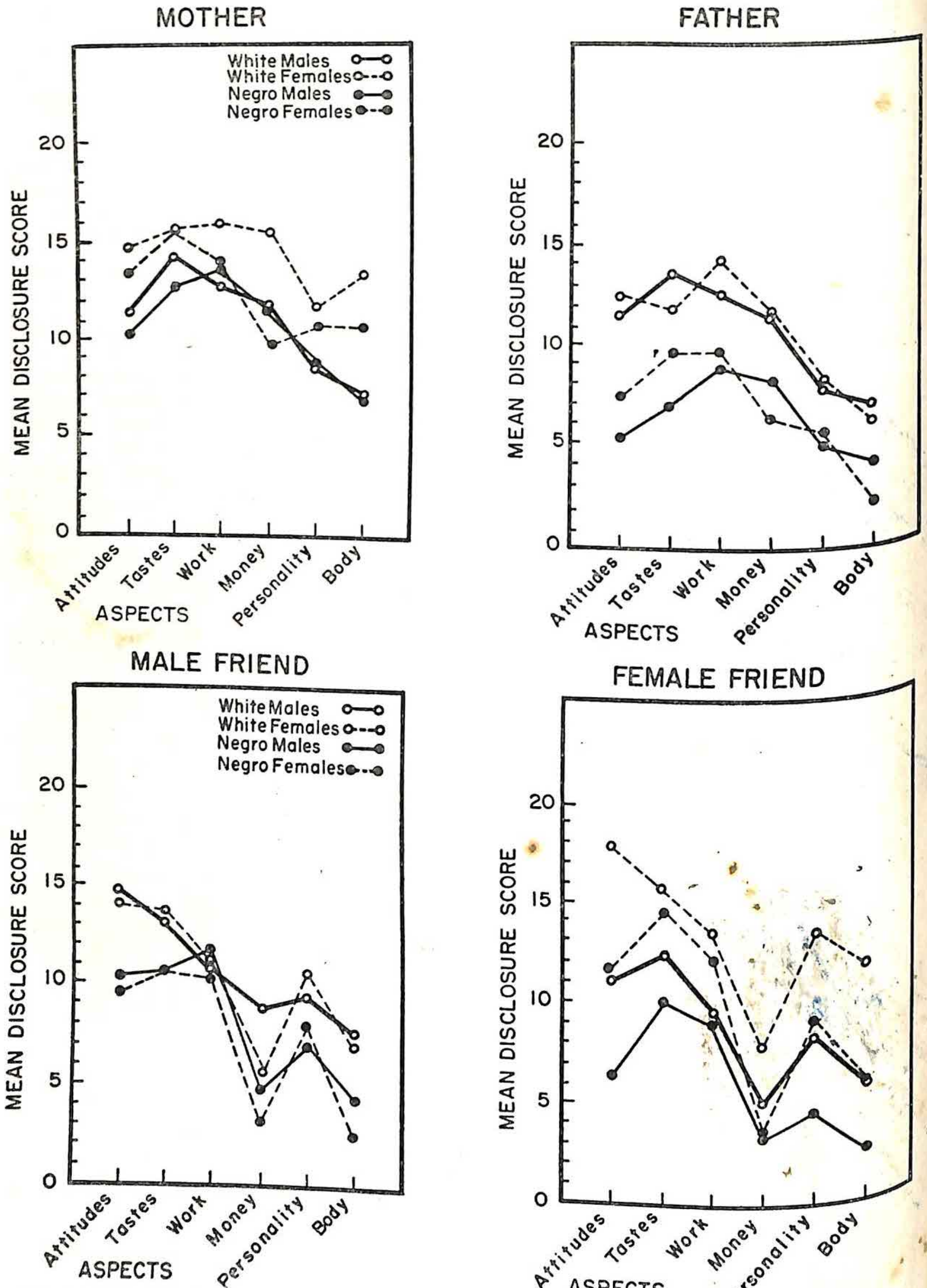


FIG. 1. MEAN SELF-DISCLOSURE OF WHITE AND NEGRO, MALE AND FEMALE *Ss* TO MOTHER, FATHER, MALE FRIEND, AND FEMALE FRIEND

(8) Type VI model for analysis of variance with mixed "between-within" effects. Critical differences for t ratios at the .01 level were computed between all groups, targets, and aspects of self when F ratios proved significant. Pearsonian r's were computed between Mother-cathexis scores and scores for disclosure to the mother, and between Father-cathexis scores and the corresponding disclosure scores, within the group of 31 nursing students to whom these two instruments had been administered.

TABLE 3

DIFFERENCES BETWEEN GROUPS IN MEAN DISCLOSURE SCORES

Group	Mean*	*SD*	Differences**		
			Negro Males	White Females	Negro Females
White males	248.50	68.60	+62.60	−45.10	+33.50
Negro males	185.90	36.10		−107.70	−19.10
White females	293.60	45.20			+78.60
Negro females	215.00	57.60			

* Highest possible score, 480.

** Critical difference for $t = 2.72$, $p < .01$, $df = 36$ is 13.95 (cf. 8, p. 93.). All differences are thus significant at $p < .01$.

TABLE 4

DIFFERENCES BETWEEN TARGETS IN MEAN DISCLOSURE SCORES

Target	Mean*	*SD*	Differences**		
			Father	Male Friend	Female Friend
Mother	72.30	19.50	+20.60***	+17.12***	+16.72***
Father	51.70	24.13		−3.48	−4.88***
Male friend	55.18	22.43			1.40
Female friend	56.58	27.70			

* Highest possible score, 120.

** Critical difference for $t = 2.63$, $p < .01$, $df = 108$ is 4.21 (cf. 8, p. 93.)

*** $p < .01$.

RESULTS

Influence of Race, Sex, Target-Differences, and Aspects of Self

Table 2 shows the results of analysis of variance of the self-disclosure scores of the 40 white and Negro *S*s, while Fig. 1 portrays mean self-disclosure scores classified by targets, groups, and aspects of self. The findings may be summarized as follows:

1. The four groups differed in total self-disclosure. Table 3 shows that the white *S*s disclosed more than the Negroes, and the females more than the males.
2. The combined group of 40 *S*s varied in amount of self-disclosure to different target-persons. They disclosed the most to Mother, and in lesser amount to Father, Male Friend, and Female Friend, as shown in Table 4. Sex- and race-differences in disclosure to the four target-persons is shown clearly in Fig. 1. Noteworthy is the consistently lower amounts of self-disclosure to Father on the part of the Negro *S*s.
3. The combined group of 40 *S*s varied in self-disclosure according to aspects of self. Table 5 shows that two clusters of aspects appeared—a "high disclosure" cluster comprised of Tastes and Interests, Attitudes and Opinions, and Work, and a "low disclosure" cluster that included Money, Personality, and Body.
4. There was significant interaction between targets and aspects, groups and targets, and groups, targets, and aspects. The group-by-aspect interaction was not significant.

TABLE 5

DIFFERENCES BETWEEN ASPECTS OF SELF IN MEAN DISCLOSURE SCORES

Aspect	Mean*	*SD*	Differences**				
			Tastes	Work	Money	Pers.	Body
Att.	45.35	14.20	−4.93***	−2.30	+13.12***	+11.30***	+19.15***
Tastes	50.28	13.98		+2.63	+18.05***	+16.23***	+24.08***
Work	47.65	13.65			+15.42***	+13.60***	+21.45***
Money	32.23	15.65				− 1.82	+ 6.03***
Pers.	34.05	13.45					+ 7.85***
Body	26.20	14.28					

* Highest possible score, 80.

** Critical Difference for $t = 2.63$, $p < .01$, $df = 180$, is 2.87 (cf. 8, p. 93).

*** $p < .01$.

Influence of Marriage

Separate analyses of variance compared married with unmarried males, and married with unmarried females. The results are shown in Table 6, for males, and Table 7, for females. In the analyses, Opposite-sex Friend and Spouse were treated as equivalent target-persons.

No differences were found between married and unmarried *S*s in total amount of self-disclosure. Figs. 2 and 3 show, however, that: (*a*) married *S*s disclosed less to the parents and the same-sex friend than unmarried *S*s, and (*b*) there was more disclosure to the Spouse than to any other target-person by married or unmarried *S*s.

Marriage thus appears to have the effect, not of increasing or decreasing the total extent to which *S*s disclose themselves, but of producing a redistribution of self-disclosure. The married *S*s "concentrated" self-disclosure

TABLE 6
ANALYSIS OF VARIANCE OF SELF-DISCLOSURE SCORES OF MARRIED AND UNMARRIED MALES

Source	Mean Square	*df*	Error Term	*F*
Between groups	37.95	1	Error (b)	.17
Error (b)	228.81	18		
Total *df* Between		19		
(Within groups comparisons)				
Between Target-Persons	232.99	3	Error$_1$ (w)	6.10**
Between Aspects of Self	477.49	5	Error$_2$ (w)	24.80***
Target × Aspect	21.72	15	Error$_3$ (w)	4.19***
Group × Target	676.95	3	Error$_1$ (w)	17.72***
Group × Aspect	1.00	5	Error$_2$ (w)	.00
Group × Target × Aspect	22.24	15	Error$_3$ (w)	4.49***
Total *df* Within		46		
Error$_1$ (w)	38.21	54		
Error$_2$ (w)	19.25	90		
Error$_3$ (w)	4.95	270		
Total Error *df*		414		
Total *df*		479		

** $p < .01$.
*** $p < .001$.

TABLE 7
ANALYSIS OF VARIANCE OF SELF-DISCLOSURE SCORES OF MARRIED AND UNMARRIED FEMALES

Source	Mean Square	*df*	Error Term	*F*
Between Groups	186.25	1	Error (b)	1.38
Error (b)	135.27	18		
Total *df* between		19		
(Within groups comparisons)				
Between Target-Persons	349.90	3	Error$_1$ (w)	6.08**
Between Aspects of Self	345.54	5	Error$_2$ (w)	18.91***
Target × Aspect	46.64	15	Error$_3$ (w)	8.00***
Group × Target	531.48	3	Error$_1$ (w)	9.23***
Group × Aspect	15.63	5	Error$_2$ (w)	.86
Group × Target × Aspect	26.01	15	Error$_3$ (w)	4.46***
Total *df* Within		46		
Error$_1$ (w)	57.56	54		
Error$_2$ (w)	18.27	90		
Error$_3$ (w)	5.83	270		
Total Error *df*		414		
Total *df*		479		

** $p < .01$.
*** $p < .001$.

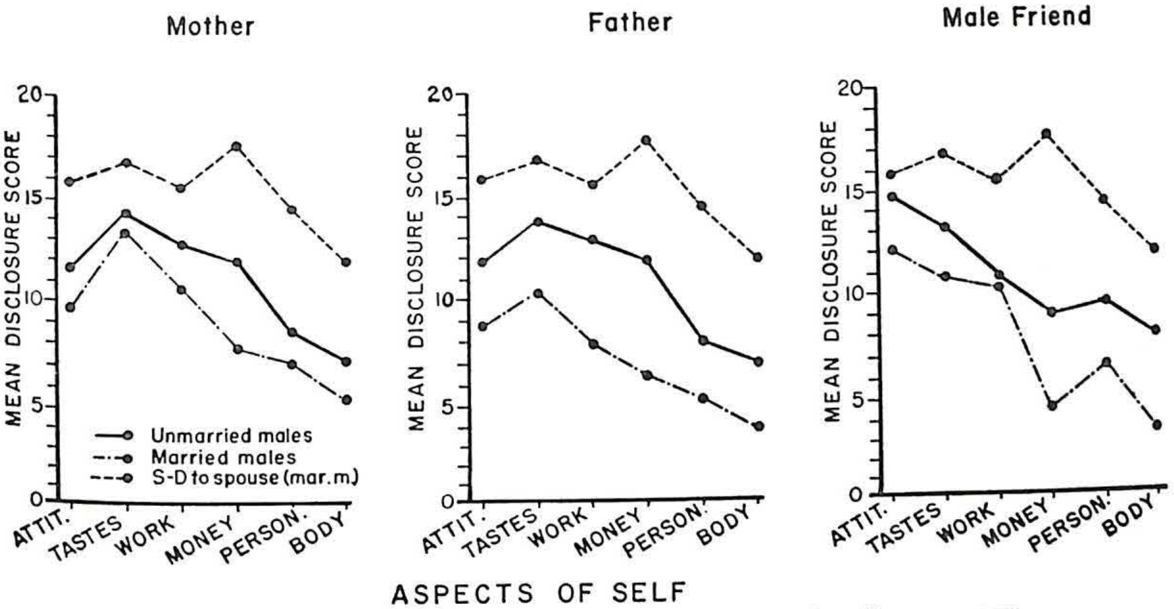


FIG. 2. COMPARISON TO MARRIED AND UNMARRIED MALES ON SELF-DISCLOSURE TO MOTHER, FATHER, AND MALE FRIEND
(Curve for disclosure to Spouse is shown in each panel)

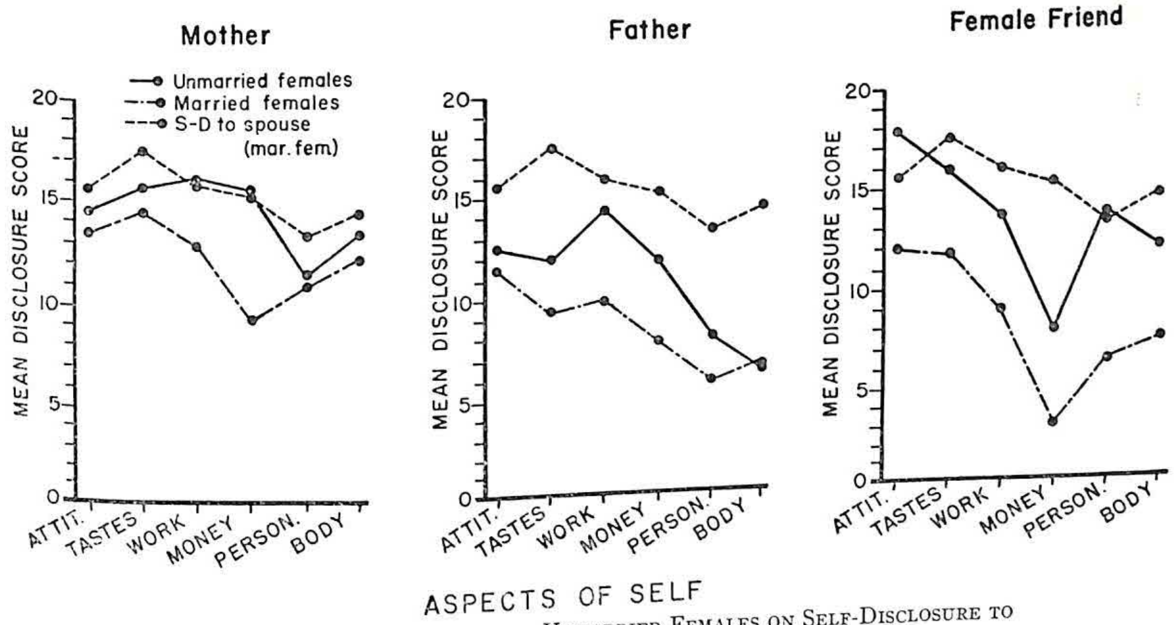


FIG. 3. COMPARISON OF MARRIED AND UNMARRIED FEMALES ON SELF-DISCLOSURE TO MOTHER, FATHER, AND FEMALE FRIEND
(Curve for disclosure to Spouse is shown in each panel)

upon the spouse, and became more reticent toward other persons. In this sense, self-disclosure enters into relations similar to those of libido in psychoanalytic theory.

Parent-Cathexis and Self-Disclosure to Parents

Total Mother-Cathexis scores for the group of 31 nursing students correlated −.63 with scores for self-disclosure to Mother; Father-Cathexis scores correlated −.53 with scores for self-disclosure to Father. Both coefficients were significant beyond the .01 level for $df = 30$. The direction of the obtained values signifies that high self-disclosure was associated with positive feelings toward the parents, while low self-disclosure was associated with attitudes of dislike toward the parents.

DISCUSSION

These preliminary findings demonstrate that self-disclosure is measurable, and that the present method for assessing it has some validity. The questions now open for exploration are virtually without limit, in view of the many possible and relevant combinations of the main factors—groups, target-persons, aspects of self, and individual differences.

Some questions suggested by the present findings may serve as guides to further exploration: Why do Negro *S*s consistently disclose less about themselves than whites, and why do females disclose more than males? Why is the mother the preferred target of self-disclosure for this age group? What is the significance of the fact that some aspects of self, for instance, Tastes and Interests, Attitudes and Opinions, and Work, are disclosed more than information about Personality, Money, and Body? Is it an artifact of the questionnaire, or does it reflect cultural consensus about what is readily disclosable and what is not? What individual traits besides feelings and attitudes toward target-persons account for individual differences in self-disclosure?

SUMMARY

A reliable questionnaire for the assessment of self-disclosure was described. Groups of both sexes, white and Negro, married and unmarried, were tested with the questionnaire for extent of self-disclosure of six different aspects of self to various target-persons—Mother, Father, Male Friend, Female Friend, and/or Spouse. The findings are summarized as follows:

1. Young unmarried *S*s, both white and Negro, showed the highest self-disclosure to Mother, with lesser amounts to Father, Male Friend, and Female Friend.

2. *S*s tended to vary the amount of self-disclosure with respect to the category of information to which an item about the self belonged. Two clusters of aspects emerged, a high disclosure cluster including Attitudes and Opinions, Tastes and Interests, and Work, and a low disclosure cluster comprised of Money, Personality, and Body.

3. White *S*s disclosed more than Negroes, and females more than males.

4. There was significant interaction among groups of *S*s, target-persons, and aspects of self.

5. Married *S*s disclosed less to Mother, Father, and Same-sex Friend than comparable unmarried *S*s. The married *S*s disclosed more to their Spouses than to any other target-persons. There was more disclosure to Spouses than to any other target-person on the part of the married and the unmarried groups.

6. A significant correlation was found between parent-cathexis and self-disclosure to the parents. The more that the parents were liked, the more disclosures were made to them.

REFERENCES

1. BLOCK, J. The assessment of communication. Role variations as a function of interactional context. *J. Pers.*, 1952, **21**, 272–286.
2. BLOCK, J., & BENNETT, LILLIAN. The assessment of communication. Perception and transmission as a function of the social situation. *Hum. Relat.*, 1955, **8**, 317–325.
3. FROMM, E. *Man for himself*. New York: Rinehart, 1947.
4. HORNEY, KAREN. *Neurosis and human growth*. New York: Norton, 1950.
5. JOURARD, S. M. Identification, parent-cathexis, and self-esteem. *J. consult. Psychol.*, 1957, **21**, 375.
6. JOURARD, S. M. *Personal adjustment: An approach through the study of healthy personality*. New York: Macmillan, 1958, in press.
7. LEWIN, K. Some social-psychological differences between the United States and Germany. In G. Lewin (Ed.), *Resolving social conflicts: Selected papers on group dynamics, 1935–1946*. New York: Harper, 1948.
8. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
9. RIESMAN, D. *The lonely crowd*. New Haven: Yale Univer. Press, 1950.

Received April 15, 1957.

TWO FORMS OF SOCIAL CONFORMITY: ACQUIESCENCE AND CONVENTIONALITY[1]

HALLA BELOFF

Queen's University, Belfast

In current psychological literature, the term conformity is applied to two distinguishable phenomena, which may be termed acquiescence and conventionality.

Acquiescence refers to the agreement with expressed group opinion in a particular experimental situation involving pressure from others. In these experiments, it is commonly arranged that the personal social influence should give rise to inconsistency with previously elicited responses, or to a seeming distortion of perceptual judgment. Studies of such dynamic conformity have been common in recent years (1, 6, 7, 9, 10, 12, 14, 15, 17, 20, 21, 22). A number of experimental devices or ruses have been employed, all of which appear to elicit a significant degree of conformity with or shift towards group opinion. These devices would seem to vary principally in the degree to which they do or do not insult the intelligence of the subjects, i.e., in the subtlety with which the pressure is applied.

By *conventionality* is meant, of course, the concurrence with the tenets, attitudes, and mores of a subject's culture or subculture. This is a very broad concept referring to the characteristics of the "insider," as he may be opposed to the individualist or eccentric. Empirical studies are fewer here, but Brodbeck *et al.* (8) have submitted the Riesman conformity typology to objective test and Bernberg (4) has related conventionality in social perception or opinion to other personality traits and to objective life criteria.

The present study was designed to obtain some objective measure of both kinds of conformity phenomena and to investigate their relationship to each other. Conventionality is here viewed as the summated end-product of past specific or piecemeal acquiescences with cultural norms. The general hypothesis, then, is that those subjects who acquiesce readily in the experimental situation are also relatively conventional; and conversely, that subjects who are independent or intransigent are also relatively unconventional. To what extent, in brief, is the use of the word conformity for both concepts justified by empirical evidence? Further relationships to three pertinent personality traits are also investigated.

METHOD

The Acquiescence Experiment

The experimental procedure followed that of Helson *et al.* (13), using the Blake simulated group technique (6).[2] Readers are referred to the original paper for a full description.

Subjects (*S*s)[3] completed a shortened version of the Thurstone-Chave War Scale (paper-and-pencil) in the usual individual way. After an interval of approximately three weeks they again responded to the Thurstone-Chave statements, under simulated group conditions; that is, they heard the stimulus statements, plus four pseudo-respondents[4] pre-recorded on sound tape, with time left between items for their spoken responses as fifth *S*. The responses of the naive *S*s were actually relayed to an adjoining room and recorded there.

The pseudo-*S*s presented strongly agreeing, neutral and strongly disagreeing backgrounds in random order, and balanced with regard to anti-, pro-, and neutral-towards-war statements.

The present situation differed from that used by Blake in one important particular. It was not practically feasible to persuade the *S*s that four others were present in other rooms concurrently. Instead, each *S* was told that the project was investigating aural attitude measurement devices and that he happened to be the fifth person to be recorded on this tape.[5] At the end of the testing period the nature and purpose of the whole experiment were fully explained to the class.

---

[1] This study was carried out while the author held a personal grant from the Medical Research Council, London.

[2] It presents a number of advantages over other systems: (*a*) it eliminates the need for the repeated presence of paid participants, (*b*) conditions within the procedure are constant from *S* to *S*, (*c*) no specially constructed apparatus is necessary. Further, the group information is presented obliquely, in contrast to those devices in which information about group decisions is presented as such, and the *S*s are invited directly to give their own responses in the light of it.

[3] A mixed class of undergraduate students attending an elementary Psychology course. For the final analysis, 35 men and 25 women, who had completed all parts of the programme, were used.

[4] These four who provided the "other *S*" stimuli were advanced students, unknown to the naive *S*s.

[5] One *S* only admitted suspecting a hoax, deduced from the fact that the other supposed *S*s, while agreeing with each other, presented such logically incon-

Acquiescence scores. A measure of acquiescence was computed by obtaining the algebraic sum of response shift on the seven-point answer scale, towards and away from the background reports. That is, the difference between the first and second response sets was considered with respect to the group report, a shift towards the background being designated positive and a shift away negative. For each *S*, the shift scores were summated for all items in which the background report had been strongly agree or strongly disagree, 12 items in all.

The range of acquiescence, or net-shift scores, was found to be −6 to +37, with a mean and mode at +11 steps (C.R. of deviation from zero = 9.18). Blake has not given a reliability coefficient in any of his reports. The present data give a Spearman-Brown corrected consistency coefficient of +0.63.

Conventionality Measures

For the purpose of empirical study, conventionality was operationally defined as the degree of agreement (expressed by a rank correlation, for example) between an individual *S* and the mean or modal response of the rest of the group. Two measures were obtained using widely disparate fields of socialized behaviour, aesthetics, and politics.

Aesthetic conventionality. The data here were the results of a simple experiment in aesthetic preference. The stimulus materials were black and white photographic slides of eleven ceramic teapots in the collection of the Victoria and Albert Museum, London. The pots ranged in time from the 17th century to the present day and included oriental as well as European products. The test was carried out as a group experiment, using the method of paired comparisons.

A rank order of preference was computed for the pots, for each *S* from his comparisons, and a mean rank order for the whole group from these. A measure of the reliability of the preferences expressed is given by the mean *zeta*, Kendall's coefficient of consistence (16, p. 123), of 0.90 for the class.

A measure of aesthetic preference or taste having been obtained, aesthetic conventionality was then measured for each *S* by the Spearman rank correlation between individual rank order and the mean rank order of the whole group (no correction being made for the inclusion of the individual's order in the group). The range of these rank correlations was found to be −.095 to +.918 (i.e., no *S* showed real opposition to class taste), with a mean at +.633.

Politicosocial conventionality. The measure again refers to the degree to which *S*s conform to the popular response of their subculture—here with regard to statements concerning a variety of religious, social, and political attitudes. Eysenck's Inventory of Social Attitudes was used as a stimulus basis. The responses to individual items provided the data for the computation of Floyd Allport's Index of Institutionalization (5; 18, pp. 82–84).

With the arrangement of a five-point multiple answer choice, a value of 2.00 for this statistic for a particular item denotes the null case, i.e., an equal distribution of checks in each category. A value of 3.00 denotes a normal distribution, i.e., some social institutionalization and conformity to it; further increases in the value of the index occur as the extent of conformity increases.

Eighteen of the 40 items manifested institutional indices above 3.00 (range 3.01 to 7.81, mean = 4.46). These items were used to compute a weighted index of political conventionality in the following manner. For each item, the weighted score consisted of the Institutional Index multiplied by the number of steps the individual's check is removed from the modal category. The item scores were summated and would in this form give an *unconventionality* score. For the sake of uniformity, this was subtracted from the maximum possible score to give a score for conventionality. The theoretically possible range is zero to 286.9, the actual scores ranging from 160.1 to 272.0, with a mean at 226.4.

RESULTS

Acquiescence and Conventionality

A simple preliminary analysis was carried out to test the hypothesis that the conformity scores of the women would be higher than those of the men (as suggested by Crutchfield, 10). Table 1 shows that this is so for acquiescence (although the difference is not significant), and very significantly so for politicosocial conventionality. In view of these sex differences, the main analysis was carried out separately for men and women. Because at least one of the variables considered does not lend itself to the usual parametric analysis (aesthetic conventionality with a markedly negatively skewed distribution), a high power nonparametric was used, the Mann-Whitney *U* test (19).

Inspection of Fig. 1 and Table 2 shows that in all cases the relationships are in the pre-

TABLE 1

SEX DIFFERENCES IN DEGREE OF CONFORMITY (Mann-Whitney *U* Test between 35 men and 25 women on the basis of a single rank order)*

Variable	Men	Women	*U*	*p*
	Mean Rank	Mean Rank		
Acquiescence	28.7	33.0	374	.17
Aesthetic conventionality	30.6	30.5	—	—
Political conventionality	22.8	36.0	245	.002

* One-tailed, as predictions were made.

Note.—For each variable in this study Rank 1 was assigned to the *lowest* score.

---

sistent responses to the attitude statements. (He was, of course, discarded from the sample.)

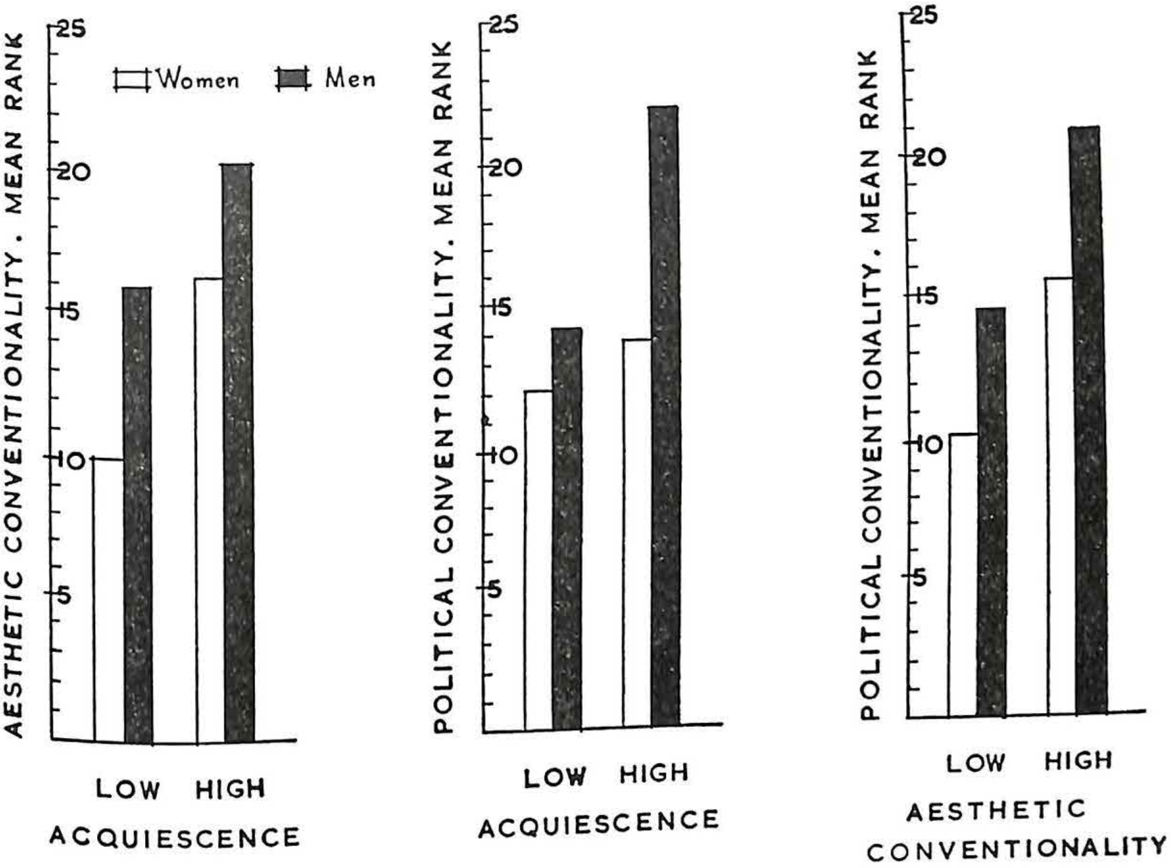


Fig. 1. Relationship Between the Three Conformity Measures
(high–low dichotomy at the median)

Note.—For the significance of the differences, see Table 2. As the number of men and women S's are not identical, mean ranks as in this figure and Table 3 are meaningfully comparable only within each sex.

dicted direction. It is seen that conventionality, in the two fields, shows the most consistent connection, reaching significance at the .05 level for both men and women.

The hypothesis concerning acquiescence and its relation to conventionality is confirmed, although the sex differences that appear complicate the conclusion. As is seen, for men the acquiescence-politics relationship is significant, and for women the relationship between acquiescence and aesthetic conventionality. It is tempting to suggest that the former field is one that is more salient for men in our culture, and the latter for women; and that the salient attitudes only, partake in the orderly integration of the *socius*. Further research, using perhaps especially selected groups of designers and students of politics would elucidate this point.

Broadly it may be concluded that acquiescence and conventionality do go together; and that this study has presented further evidence therefore for a generalized tendency towards and away from conformity.

TABLE 2

Relationship between the Three Measures of Conformity

Variables	Men		Women	
	U	p*	U	p*
Acquiescence & aesthetic conventionality	115	.11	39	.02
Acquiescence & political conventionality	85	.02	67	.27
Aesthetic & political conventionality	96	.03	45	.05

* One-tailed.

Personality Test Correlates

The three personality traits considered here were chosen both for their seemingly obvious

TABLE 3

RELATIONSHIP BETWEEN CONFORMITY MEASURES (MEDIAN SPLITS) AND PERSONALITY VARIABLES (MEAN RANKS)

Personality Trait	Men				Women			
	Low	High	*U*	*p**	Low	High	*U*	*p**
Acquiescence								
Ascendance	20.1	15.7	114	.10	11.7	14.4	61	.18
Neuroticism	*15.0*	*20.8*	*103*	*.05*	*15.7*	*10.0*	*42*	*.03*
Calif. F score	*15.4*	*20.5*	*108*	*.07*	*11.0*	*15.1*	*52*	*.08*
Aesthetic Conventionality								
Ascendance	18.8	17.1	137	.31	11.3	14.8	56	.12
Neuroticism	21.1	14.3	93	.03	15.2	10.6	49	.06
Calif. F score	15.7	20.6	110	.08	12.5	13.5	72	.37
Political Conventionality								
Ascendance	18.3	17.7	149	.42	10.7	15.4	50	.06
Neuroticism	21.0	14.8	99	.04	13.4	12.6	73	.39
Calif. F score	15.6	20.2	113	.09	12.3	13.7	70	.32

* One-tailed.

relevance, and because although they have been the subject of previous investigation, findings with regard to them have by no means been consistent.

1. *Ascendance-submission.* It has been shown (13) that male *S*s who acquiesce in a simulated group situation have significantly lower ascendance scores on the A-S Reaction Study inventory. This is hardly a surprising finding, in that any of the acquiescence devices might well have been designed just as an objective validating experiment for the inventory.

From Table 3 it can be seen that the male *S*s here do display a tendency in this direction; higher scores on all three forms of conformity going with lower ascendance scores. The women, however, manifest a consistent trend in the opposite direction, the more conforming *S*s being the *more* ascendant ones.

2. *Neuroticism.* The consensus of opinion with regard to neuroticism seems to be that although independent nonconformists admit to having more neurotic tendencies when questioned more or less directly (2, 9, 15), it is the conformers who are perhaps *really* more neurotic (10, 14). Using the Maudsley Personality Inventory designed by C. M. Fransk (11), it was predicted that this simple and rather obvious self-rating procedure would show the conformers to be less neurotic.

Table 2 shows that again the results are by no means simple. For conventionality, those data that reach significance or approach it, do confirm the prediction, i.e., the unconventional *S*s show higher neuroticism scores. Considering acquiescence however, the women too confirm the hypothesis, while the men attain a result of equivalent significance in the opposed direction.

3. *Authoritarianism.* One of the assumptions both overt and covert in the Californian writings on the authoritarian personality is that authoritarian persons are conformers. And indeed this is the one consistent relationship that has been demonstrated in all the studies concerned with it. Some report correlations between acquiescence and actual scores on the California E or F scales (7, 10, 20); others with ratings, etc., of traits underlying the authoritarian concept (2, 14).

Using the F scale (Forms 45 and 40), that connection is confirmed here, albeit only at the .07 and .08 confidence levels for men and women respectively. With regard to conventionality only the men's results approach significance.

However, before accepting a simple link between the tendency towards authoritarianism and conformity, as has been done so far by other writers, it is necessary to consider the school of thought (3) which has considered that a large part of the variance of the F scale in its present form is taken up by some behaviour which is best called acquiescence! In the light of the evidence at present available, the most economic explanation of the acquiescence-F score relationship found here and in previous studies must be that the acquiescence-inducing situations and the orthodox F scale just measure the same thing.

Obviously, further investigation must now concern itself not with the development of yet more acquiescence devices but with the clarification of the place of conformity in general personality organization and, more particular, with the meaning and significance that conformity has for men and women, in order to elucidate the part that conformity plays in their adjustment to their respective roles. This meaning or significance would appear to be very different for the two sexes. As we have seen in our data, acquiescent men appear to be more submissive and more neurotic, acquiescent women more ascendant and more stable.

SUMMARY AND CONCLUSION

In this project, the two principal aspects of conformity were empirically studied; i.e., the dynamic process of immediate conformity, termed acquiescence; and long term results of conformity to cultural norms, termed conventionality.

The relationship between the two phenomena was studied, as were correlates with three seemingly pertinent personality characteristics. The main hypothesis put forward was that persons scoring relatively high on the acquiescence measure would tend also to score high on conventionality indices.

On the basis of performance in a Blake-type simulated group situation, a group of 60 undergraduate university students were divided into two categories of high and low acquiescence. Conventionality, operationally defined as high agreement between an individual's response and the mean or modal response of his group, was measured in the areas of aesthetics and politics. Photographs of teapots and the Eysenck Social Attitudes Inventory provided stimulus material for the computation of conventionality scores by means of rank correlations and the Allport Institutional Index.

Analysis by means of the Mann-Whitney *U* test confirmed the hypothesis of a positive relationship between all three conformity scores. Certain sex differences appeared: women *S*s tended to exhibit a greater degree of conformity than men; and also, the acquiescence-conventionality relationship reached significance for men in politics and for women in aesthetics.

The literature had indicated that present knowledge of conformity-personality correlates is in a confused state. The present data confirm this point. Striking sex differences again add complexity to the results. The only relationship that is found to be uniform for both sexes and is congruent with past reports is that between acquiescence and high scores on the California F Scale. The possibility that this is the result of an experimental artifact in the present form of the F scale is discussed.

REFERENCES

1. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
2. BARRON, F. Some personality correlates of independence of judgment. *J. Pers.*, 1953, **21**, 287–297.
3. BASS, B. M. Authoritarianism and acquiescence. *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
4. BERNBERG, R. E. Personality correlates of social conformity. *J. appl. Psychol.*, 1954, **38**, 148–149.
5. BERNBERG, R. E. A measure of social conformity. *J. Psychol.*, 1955, **39**, 89–96.
6. BLAKE, R. R., & BREHM, J. W. The use of tape recording to simulate a group atmosphere. *J. abnorm. soc. Psychol.*, 1954, **49**, 311–313.
7. BLOCK, JEANNE, & BLOCK, J. An interpersonal experiment on reactions to authority. *Hum. Relat.*, 1952, **5**, 91–98.
8. BRODBECK, A. J., NOGEE, P., & DI MASCIO, A. Two kinds of conformity: A study of the Riesman typology applied to standards of parental discipline. *J. Psychol.*, 1956, **41**, 23–45.
9. CERVIN, V. Experimental investigation of behaviour in social situations: I. Behaviour under opposition. *Canad. J. Psychol.*, 1955, **9**, 107–116.
10. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
11. EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, in press.
12. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.
13. HELSON, H., BLAKE, R. R., MOUTON, JANE S., & OLMSTEAD, J. A. Attitudes as adjustments to stimulus, background and residual factors. *J. abnorm. soc. Psychol.*, 1956, **52**, 314–322.
14. HOFFMAN, M. L. Some psychodynamic factors in compulsive conformity. *J. abnorm. soc. Psychol.*, 1953, **48**, 383–393.
15. JANIS, I. L. Anxiety indices related to susceptibility to persuasion. *J. abnorm. soc. Psychol.*, 1955, **51**, 663–667.
16. KENDALL, M. G. *Rank correlation methods.* London: Griffin, 1948.

17. OLMSTEAD, J. A., & BLAKE, R. R. The use of simulated groups to produce modifications in judgment. *J. Pers.*, 1955, **23**, 335–345.
18. PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
19. SIEGEL, S. *Non-parametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.
20. WELLS, W. D., WEINERT, G., & RUBEL, MARILYN. Conformity pressure and authoritarian personality. *J. Psychol.*, 1956, **42**, 133–136.
21. WIENER, M., CARPENTER, JANETH T., & CARPENTER, B. External validation of a measure of conformity behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 421–422.
22. YOUNG, N., & GAIER, E. L. A preliminary investigation into the prediction of suggestibility from selected personality variables. *J. soc. Psychol.*, 1953, **37**, 53-60.

Received April 17, 1957.

THE ROLE OF MOTIVATION IN PSYCHOLOGICAL STRESS[1]

WILLIAM VOGEL, ROBERT W. BAKER, AND RICHARD S. LAZARUS[2]

Clark University

LAZARUS and Baker (5) recently have presented a theoretical and methodological framework for the study of behavior under stress. Their conception is that the motivational state of the individual in relation to the stressor conditions imposed upon him determines whether or not a state of stress occurs. Experimentation within this framework thus involves the systematic use of motivation as an independent variable. The purpose of this paper is to present some evidence concerning the role of motivation in determining response to stressor conditions.

In our experimental studies we distinguish between *induced* and *intrinsic* motivation. By induced motivation we mean those transitory organismic states which we expect to arouse or instil through the use of ego-involving instructions or through the manipulation of incentives. Intrinsic motivation, on the other hand, refers to motive states that are presumed to be relatively enduring characteristics of individuals. These are typically inferred from apposite behavioral samples, including fantasy productions (cf. McClelland, 7).

A further distinction may be made between kinds of behavior samples from which intrinsic motivational states are inferred. On the one hand, "long run, everyday" behavioral characteristics of individuals may be used, as when achievement motivation is inferred from the number of hours that a student studies per week. On the other hand, imaginational production or fantasy may be employed, as McClelland and his associates have done in the case of achievement (7), affiliation (1, 10), and power motivation (13). The assumption is often made that the incidence of need related imagery in the stories a person tells to picture stimuli (especially in the absence of external achievement cues) reflects a relatively stable and persistent motivational characteristic.

All these means of studying motivation are relevant to the three theoretical issues to which the present experiment is addressed. The first of these issues involves the relationship between a stressor condition[3] and motivational characteristics of individuals. If there is no relevant motive state to be thwarted, then according to our theoretical position there is no stress; if there is a strong motive state, but it is not engaged by the stressor condition, then there is likewise no stress; if there is a motive state, and it is engaged by the stressor condition, the resultant stress state and its consequences vary as a function of the intensity of the motive state. In this study, we focus upon the last of these alternatives and investigate behavioral output and affective arousal as a function of induced motive states varying in intensity. If differential achievement motivation is induced in two groups of subjects, we would expect that failure-information (the stressor condition) should be more thwarting to the more highly motivated subjects, resulting in greater affective arousal (i.e., heightened physiological reactivity) and a greater increase in effort on the task.

The second problem addressed by our experiment is the generality of motive states, i.e., the extent to which a motivational characteristic inferred from one behavioral context is also operative in another. By definition, this problem refers to intrinsic rather than induced motivation. The study seeks to determine whether degree of motivation inferred from level of academic effort and accomplishment in college is associated with behavioral output on a performance task and accompanying physiological reactivity under various experimental conditions, including a stressor. The experimental design also permits the effect of intrinsic motivation on behavioral output and physiological reactivity to be compared with those of induced motivation.

The third issue to be considered concerns the relationship between the expression of a motive in imaginational productions or fantasy and its actualization in behavior. Fantasy has been used as a means of measuring intrinsic motivation on the assumption that this relationship is always direct and positive (7), but evidence has recently been presented that this assumption is not entirely warranted (6). The present experiment investigates the relation of a fantasy measure of need for achievement (7) to behavioral output and physiological reactivity under various experimental conditions, and also compares this measure with indices derived from other behavioral sources.

---

[1] This investigation was supported by research grant (M-734) from the National Institute of Mental Health, Public Health Service.

[2] Now at the University of California at Berkeley.

[3] By stressor conditions we mean those experimental conditions which are intended by the investigator to be stressful, but which are not necessarily responded to as such by the *S*. The term "stress" we use to refer to a state of the organism, not to a class of conditions.

METHOD

Subjects and Procedure

Forty male undergraduate college students (mostly freshmen) who volunteered for the study were selectively assigned to two groups of 20 each in such a manner that the two groups represented similar distributions of scores on two measures of intrinsic motivation, McClelland's need Achievement score and a measure relating academic accomplishment to intellectual ability. One group was told that they were going to perform a task that would allow the experimenter to evaluate the effect of work on physiological activity. This information was intended to be relatively less motivating or ego-involving than the treatment given the second group, which was informed that their performance on the task would provide the experimenter with a means of assessing their potential for high level academic and vocational functioning. All of the *S*s in the experiment were tested individually.

Before the differentially motivating instructions were given, the two groups were given a "practice" session on the experimental task, ostensibly in order to become familiar with it. Actually, this session ensured the comparability of the two groups regarding physiological reactivity, as well as affording a baseline for gauging the effects of conditions. The practice condition lasted approximately four minutes during which all *S*s performed an equal number of task items.

The experimental task was the McKinney Reporting Test (8), which presents a long series of circles containing three symbols in various quantities. The *S*'s task is to count as quickly and accurately as possible the number of symbols of each kind which each circle contains.

Following the practice period, the differentially motivating instructions were given and a 12-minute "work" condition ensued. At the end of this period, *S*s were informed that their performance was inadequate and that they would be given an opportunity to do better. During a second ensuing 12-minute period of performance (the failure-stressor condition) both groups of *S*s were given false, unattainable norms every minute to induce a sense of failure.

Following the failure-stressor condition, *S*s were administered cards 6BM and 7BM of the Thematic Apperception Test and the machineshop picture of the McClelland series (7), which they had taken under neutral conditions on an earlier day. If a different set had been used the second time following McClelland's procedure, it would not have been clear whether obtained differences in imagery were a function of conditions or of differential achievement cues of the cards. McClelland's attempt to equate the cue value of pictures is of questionable value in studies of individual differences, because the cue value of a picture may vary for different individuals and *S* populations.

At the same earlier session, the Taylor Anxiety scale (12) was also administered, and *S*s were asked to indicate on a chart their activities for the previous week hour by hour to provide reliable information about their study habits.

Physiological Measures

Measures of physiological reactivity were taken as follows. In the practice condition, galvanic skin response (GSR) readings were made during the first and third minutes, pulse was recorded at two minutes, and diastolic and systolic blood pressure at four minutes. The same procedure was followed in the two subsequent conditions, being repeated three times during each 12-minute period.

The GSR apparatus used was a type 310-A Z-Angle meter, manufactured by The Technology Instrument Corporation of Acton, Mass. An internal oscillator was used, operating with 400 cps sources of test signal. Approximately .04 volts of AC current was impressed across the unknown impedance. Two electrodes made from dimes were placed, after electrode paste had been applied, to the first and index fingers of the *S*'s left hand. Direct impedance measurements were recorded.

The GSR reading employed was the average of three readings taken at 20-second intervals over a one-minute span. The pulse recorded was based on a constant reading taken over a one-minute span. Blood pressure was taken by the usual auscultatory method with a standard mercury sphygmomanometer. Diastolic blood pressure was read at the point of disappearance of the sound. For both diastolic and systolic blood pressure, two readings were taken per minute, and the average of the two was recorded as the reading for that minute. Of the three diastolic pressures, three systolic pressures, and three pulse readings obtained for each condition, the peak response for each measure was retained for further analysis. Of the six GSR responses recorded for each condition, the lowest impedance reading was retained for further analysis. All impedance readings were converted into conductance scores.

The index of physiological reactivity employed was adapted from Lacey (3). The measures of systolic and diastolic blood pressure, pulse, and GSR were separately normalized and converted into *T* scores. The most reactive measure for each *S* across the several conditions was then determined by adding *T* scores across conditions. The channel for each individual having the highest *T*-score total was selected as his most reactive channel, and his index of physiological reactivity was based on data from this channel. This index is expressed as a logarithmic transformation of percentage change from the practice condition to the work and stressor conditions respectively, thus correcting for base level.

Lacey's method has its drawbacks, as the most reactive measure over the course of the experiment might not be most reactive in a particular condition. However, 66 per cent of the *S*s showed perfect consistency through all conditions, and another 21 per cent were highly consistent in that measurements in their most reactive channel did not differ by more than 2 *T*-score units from their most reactive measure in any single condition.

There are essentially two alternatives to Lacey's technique. One is an index based on an average of scores from all autonomic channels. This method involves the questionable assumption that a person's somatic reactivity is a direct function of the number of channels in which he expresses this reactivity, although it avoids the assumption that the highest physiological response is the best measure of arousal. The authors obtained very few significant results whenever it was applied.

The other alternative is to utilize the *S*'s most reactive measure in *each* condition. This method, which is consistent with an assumption of extreme specificity, has the drawback of giving maximal weight to chance autonomic fluctuations. A larger number of uninterpretable results were derived with this method than with any other. Accordingly, Lacey's method seemed the best for the present study. Many of these problems in the use of autonomic measures in psychological research are discussed by Lacey in a recent article (4).

RESULTS AND DISCUSSION

Relationship Between a Stressor Condition and Motive States

Table 1 shows change in physiological reactivity from the baseline measures of the "practice" condition under the "work" condition, after the different instructions had been given the *S*s; and under the stressor condition, when both groups of differentially motivated *S*s were given failure-information.

It will be noticed that in the work condition, the more strongly motivated of the two groups shows the greater increase in physiological reactivity. However, and this we did not anticipate, physiological reactivity in the two groups is at the same level under the stressor condition.

Table 1 also shows that similar results are obtained with performance output on the McKinney Reporting Test, i.e., number of items attempted.[4] In the work condition, the strongly motivated group attempts significantly more items than the weakly motivated group, while in the stressor condition both groups increase in output, the *greater* increase in the weakly motivated group rendering the two groups again comparable.

TABLE 1

PHYSIOLOGICAL REACTIVITY AND TASK PERFORMANCE AS FUNCTIONS OF INDUCED MOTIVATION AND EXPERIMENTAL CONDITIONS

Groups and Conditions	*T* Scores of Logarithm Percent Physiological Change*	Items Attempted McKinnney Test**
High induced motivation ($N = 20$)		
Work	60.42	33.04
Failure-stressor	65.43	40.89
Low induced motivation ($N = 20$)		
Work	52.29	27.89
Failure-stressor	60.82	40.09

* Interaction significant below the .05 level ($F = 4.58$).

** Interaction significant below the .01 level ($F = 8.98$).

The findings presented in Table 1 do not support our hypothesis that different levels of induced motivation are associated with differential response to a stressor condition. Our first inclination was to interpret these results as suggesting that for the relatively unmotivated *S*s, failure-information had an effect equivalent to that of the motivating instructions for the other group, making level of physiological reactivity and performance output comparable for the two groups. Our second inclination was to look for interactive effects which, as we had learned in previous studies (5), could obscure real differences. Accordingly, we examined the relationship between level of induced motivation, performance output, and physiological reactivity under the stressor condition.

We first determined that there is no relationship between physiological reactivity and performance output under the stressor condition when induced motive state is not taken into account. That is, *S*s with marked physiological reactivity are just as likely to have low as high behavioral output. However, when these two variables are considered in their relationship to level of induced motivation, a strong interaction appears.

Table 2 shows number of items attempted on the McKinney Reporting Test under the stressor condition as a function of level of induced motivation and degree of physiological reactivity. It is clear from this table that strongly motivated *S*s who were most reactive physiologically attempted signifi-

---

[4] Errors do not account for an appreciable amount of variation and are therefore not presented as a dependent variable; number of items attempted and number of items correct correlate very highly.

TABLE 2

TASK PERFORMANCE UNDER FAILURE AND MAS SCORE AS FUNCTIONS OF INDUCED MOTIVATION AND LEVEL OF PHYSIOLOGICAL REACTIVITY

(N = 6 in each group)

Groups	Items Attempted, McKinney Test*	MAS Score**
High induced motivation		
Somatically reactive	43.39	10.0
Somatically nonreactive	34.00	14.5
Low induced motivation		
Somatically reactive	35.44	16.2
Somatically nonreactive	42.72	12.5

* Interaction significant below the .01 level (F = 14.24).

** Interaction nonsignificant (F = 2.60; that required for significance at the .05 level = 4.35).

cantly more items than strongly motivated *S*s who were least reactive. On the other hand, *weakly* motivated *S*s who were most reactive physiologically attempted significantly *fewer* items than weakly motivated *S*s who were least reactive. In other words, under strong motivation the relationship between physiological reactivity and performance output is *positive*, while under weak motivation it is *negative*. These findings with respect to induced motivation had no parallel with respect to intrinsic motivation.

Why should physiological reactivity be positively related to performance under strong induced motivation, and negatively under weak motivation? It seems appropriate that highly motivated *S*s, told that performance was very important, should react to failure-information with relatively greater physiological response and that less motivated *S*s, given to believe that performance was of no great consequence, should react with relatively little physiological response. By the same token, it seems *in*appropriate that highly motivated *S*s should show relatively little physiological response to failure-information, or that weakly motivated *S*s should show relatively greater reactivity.

Interestingly, those individuals responding inappropriately—highly motivated-physiologically unreactive; weakly motivated-physiologically reactive—show less performance output on the experimental task than those who respond appropriately under the stressor condition. Possibly the inappropriate responders, being also less productive, may be the less well-adjusted *S*s. Since the Taylor Manifest Anxiety Scale (MAS) has been demonstrated to correlate highly with a scale of neuroticism (14), this hypothesis can be checked; the relevant data are in the second column of Table 2. The differences are not statistically significant, but there is a suggestive tendency for the inappropriate responders to have higher scores.

The Generality of an Intrinsic Motive State

The availability for our *S*s of academic records and a measure of intellectual ability made it possible to devise a motivational measure in terms of the relationship between accomplishment and capacity. "Overachievers" and "underachievers" were identified by subtracting each student's score (expressed in standard score form) on the American Council on Education Psychological Examination for College Freshmen (administered in all cases just prior to his matriculation as a freshman) from his college grade point average (also expressed in standard score form).

We assumed that overachievers tend to be highly motivated to achieve and underachievers poorly motivated. In present terminology, the difference, if it be motivational, clearly pertains to intrinsic motivation. The motivational interpretation of over- and underachievement is substantiated by the measure of study habits. Overachievers reported an average of 34 hours spent studying for the week, while underachievers reported an average of 14 hours.

We used this index of intrinsic motivation as an independent variable in the prediction of behavior in the experimental situation, expecting that overachievers would attempt more items and show more accompanying physiological activity. The data for this comparison are presented in Table 3. It is clear that our expectations were not realized. The underachievers have greater performance output and are more physiologically reactive than the overachievers.

We would account for this apparent paradox by suggesting that we attributed unwarranted generality to the intrinsic motive state, neglecting the important interactive relationship that may be presumed to exist between a motive state and the operations required by a task. We failed to take into account the

TABLE 3

Number of Items Attempted on the McKinney Reporting Test and Somatic Reactivity as a Function of Intrinsic Motivation

Groups (N = 20 in each case)	Items Attempted, McKinney Test**	Somatic Reactivity as T Scores of Logarithm Percent Physiological Change*
Overachievers	31.98	58.07
Underachievers	39.02	62.46

Note.—Performance data for both groups have been combined across all conditions since there were no significant interactions with conditions; physiological data were combined across the work and stressor conditions only, since they are expressed as log percentage changes from the practice condition.

* Difference significant below the .05 level (F = 4.47).

** Difference significant below the .01 level (F = 10.39).

differences between academic tasks and the McKinney test, the former requiring conceptual operations, the latter being a simple, repetitive, monotonous, perceptual-motor task. After the fact, it seems plausible that the overachievers, being motivated to achieve in conceptual tasks, were "left cold" by the McKinney test; i.e., the test failed to engage the achievement motivation which they had demonstrated in another context, while it did engage the achievement motivation of the underachievers by virtue of its perceptual-motor character. Unpublished questionnaire observations by the authors have suggested that underachievers had greater interest in motoric tasks, such as athletics, while overachievers spent more time at, and expressed more interest in, academic and intellectual activity. This explanation remains to be tested. The implication that motive states should not be considered as independent of operations required by tasks, in view of individual differences in predilection for different operations, nevertheless now seems to us almost self-evident.

The Relationship Between Expression of a a Motive in Imaginational Productions and its Actualization in Behavior

In earlier studies (6) we used as a measure of intrinsic motivation, a method devised by McClelland and his associates (7) for inferring achievement motivation from imaginational productions to picture stimuli. However, results of such studies led us (6) to question two assumptions underlying the use of a fantasy index of motivation: (*a*) that need states are necessarily expressed in fantasy, and (*b*) that when they are, there is necessarily a direct and positive relationship between the amount of need-related imagery and the degree of actualization of the need in behavior. If a person shows striving for achievement in his everyday behavior, one might expect him also to express this need in his imaginational productions, and vice versa. If fantasy can also be conceived as serving as a substitutive activity in the absence of need gratification, however, the expression of need in fantasy need not correspond with its expression in other forms of behavior. Lazarus *et al.* (6) have advanced the view that whether a need is expressed in fantasy, in motoric activity, or both depends in large part upon regulative processes in the personality.

With the present data, the need Achievement score was examined in relation to performance output and physiological reactivity. If it is a "pure" index of achievement motivation, it might be expected to be positively associated with number of items attempted on the McKinney Reporting Test and with the index of physiological reactivity. Two sets of need Achievement scores are available for our *S*s: those derived from story-productions taken under innocuous conditions on the "control" day (hereafter referred to as "neutral"), and those derived on the "experimental" day, following an experience of failure (hereafter referred to as "aroused"). For purposes of analysis of variance, *S*s were ordered separately according to their neutral and their aroused need Achievement scores and divided in each case at the median.

No significant relationship with either dependent variable was found for the need Achievement measure taken under the neutral condition. Thus, when *S*s were administered the picture stimuli in the absence of external achievement cues, the amount of need-related imagery expressed does not predict behavior in the experimental context. In addition, there were no significant interactions between neutral or aroused need Achievement measures and the experimental conditions, including the stressor. Neither were there significant relationships between the need Achievement score and induced motivation: The highly motivated group did not have higher arousal scores than the weakly motivated group.

TABLE 4

NUMBER OF ITEMS ATTEMPTED ON THE MCKINNEY REPORTING TEST AS A FUNCTION OF AROUSED NEED ACHIEVEMENT

Groups (N = 20 in each case)	Items Attempted, McKinney Test*
High n Achievement	32.74
Low n Achievement	38.51

* Difference is significant below the .05 level (F = 5.71).

There was, however, a main effect for the aroused need Achievement measure, but it was in the direction opposite to what would have been predicted on the basis of an assumed direct and positive relationship between the fantasy and behavioral expression of needs in performance output. Table 4 shows that, independent of conditions, persons with relatively greater expression of achievement-related imagery after being given failure information attempt *fewer* items on the experimental task than persons with lesser expression of such imagery.

This finding is comparable to those cited in the preceding section, where the over- and underachievement index was found to be negatively related to performance and physiological reactivity. In other words, here again one might argue that the McKinney Reporting Test does not challenge the *S*s with high achievement fantasy scores. While these results together suggest the possibility of a positive relationship between the need Achievement score and the over- and underachievement index, neither analysis of variance nor correlational methods reveal any such association. In relation to our negative relationship between need Achievement and performance on the McKinney Reporting Test, Atkinson and Reitman (2) have found no relationship between need Achievement and performance on a perceptual-motor task (drawing X's in circles), and a positive relationship between need Achievement (under neutral conditions) and performance on an arithmetic task. We have suggested earlier that a perceptual-motor task was not involving for the high need Achievers, while a conceptual task might be. The Atkinson and Reitman finding adds credence to this possibility, and also suggests that the nature of the need Achievement-performance relationships depends upon experimental conditions.

TABLE 5

PHYSIOLOGICAL REACTIVITY AND MAS SCORE AS FUNCTION OF NEUTRAL AND AROUSED NEED ACHIEVEMENT AND EXPERIMENTAL CONDITIONS (N = 10 in each cell)

Achievement Imagery	*T* Scores of Logarithm Percent Physiological Change**		MAS Score
	Work	Failure-stressor	
High Imagery, neutral condition			
High imagery, aroused condition	56.04	65.12	21.5
Low imagery, aroused condition	56.82	61.16	11.3
Low Imagery, neutral condition			
High imagery, aroused condition	56.12	59.99	12.40
Low imagery, aroused condition	60.01	69.30	13.20

* Double interaction significant below the .05 level (F = 4.87).

** Triple interaction significant below the .01 level (F = 8.77).

In Table 5, evidence for a triple interaction between neutral and aroused need Achievement imagery and experimental conditions, including the failure stressor, is presented. According to this table, the most physiologically reactive group of *S*s, and the one also showing the most change in physiological activity from the work to the failure stressor condition, is the group which has the lowest need Achievement scores under both the neutral and aroused conditions. This finding is comparable to one by Raphaelson (9) who found that *S*s with low need Achievement scores (under aroused conditions) showed greater physiological reactivity (GSR) in a failure situation than high need Achievers. Raphaelson interpreted his results in terms of "fear motivating avoidance" on the part of the low achievement group, while we propose that there is more "positive interest" in a perceptual-motor task on the part of the low need Achievement *S*s, and less involvement on the part of the high need Achievers. Either interpretation would account for the negative relationship between need Achievement score and physiological reactivity. However, the additional finding, presented here, that there is also an inverse relationship between need Achievement and performance output, seems to support the notion that low need Achievers are more ego-involved in perceptual-motor tasks while high need Achievers have greater involvement in abstract-conceptual tasks.

As a final comparison, we analyzed the relationship between the need Achievement scores under the neutral and aroused conditions and the Taylor Anxiety scale (12), obtaining the significant interaction also shown in Table 5. *S*s with high need Achievement scores under both neutral and aroused conditions are seen to have the highest anxiety scores. In conjunction with the negative relationships between need Achievement scores and physiological reactivity and performance output, this relationship between need Achievement and anxiety score is difficult to interpret in terms of drive level as might be suggested by Taylor and others (11). Rather, we are inclined to believe that *S*s giving high need Achievement scores, even in the absence of clear, external achievement cues (the neutral condition), give more evidence of neurotic disturbance. The foregoing results strongly suggest that the fantasy index should not be regarded as a pure measure of motivation that is actualized in behavior.

SUMMARY AND CONCLUSIONS

This paper reports an experiment designed to investigate motivational characteristics of individuals as a source of individual differences in response to stressor conditions. Motive states were manipulated in two essential ways: (*a*) through induction or arousal by use of ego-involving instructions (induced motivation) and (*b*) through inference from fantasy productions and academic accomplishment as well as study habits (intrinsic motivation). Effects of such motive states were sought in two kinds of behavior under varied experimental conditions, including a failure-stressor. The two kinds of behavior examined were performance output on a perceptual-motor task (the McKinney Reporting Test) and accompanying physiological reactivity.

The study is addressed to three theoretical issues. The first of these refers to the relationship between a stressor condition and induced motive states which vary in intensity. Our expectation that there would be a simple relationship between strength of induced motive state and behavior under a stressor condition was not borne out. However, different motive states are found to be associated with differential response to a stressor *when motivation is examined in interaction with level of physiological reactivity*. An attempt was made to account for the specific interactive effects in terms of degree of appropriateness of the reaction of the subject to the conditions.

The second theoretical issue involves the generality of an intrinsic motivational state inferred from the relation between academic accomplishment and intellectual capacity. Our results indicate that strength of this motive state is inversely related to performance output and physiological reactivity in the experimental situation. We attempt to account for this seeming paradox by suggesting that a motivational characteristic inferred from one behavioral context (i.e., the academic situation) is not necessarily operative in a different context (i.e., the McKinney Reporting Test).

The third theoretical issue of relevance in this experiment pertains to the relationship between the expression of a motive in imaginational productions or fantasy and its actualization in behavior. A positive and direct relationship is often assumed. That the relationship is not always positive and direct is shown by our findings, where amount of achievement-related imagery is inversely related to motivated behavior. We suggest that whether a need is expressed in fantasy, in behavior, or both will be a function of regulative processes in the personality.

The findings of this experiment indicate the complexity of the role of motivation in determining response to stressor conditions and highlight the methodological difficulties involved in the systematic investigation of motivational variables. At the same time, the results indicate the importance of investigating motivational variables in stress experimentation.

REFERENCES

1. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.
2. ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–367.
3. LACEY, J. I. Individual differences in somatic response patterns. *J. comp. physiol. Psychol.*, 1950, **43**, 338–350.
4. LACEY, J. I. The evaluation of autonomic re-

sponses: toward a general solution. *Ann. N. Y. Acad. Sci.*, 1956, **67**, 123–164.

5. Lazarus, R. S., & Baker, R. W. Personality and psychological stress—A theoretical and methodological framework. *Psychol. Newsltr.*, N.Y.U., 1956, **8**, 21–32.
6. Lazarus, R. S., Baker, R. W., Broverman, D. M., & Mayer, J. Personality and psychological stress. *J. Pers.*, in press.
7. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
8. McKinney, F., Strother, G. B., Hines, R. R., & Allee, R. A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.
9. Raphaelson, A. C. The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13–19.
10. Shipley, T. E., & Veroff, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.
11. Spence, K. W., & Taylor, Janet. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
12. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
13. Veroff, J. Development and validation of a projective measure of power motivation. *J. abnorm. soc. Psychol.*, 1957, **54**, 1–8.
14. Winne, J. F. A scale of neuroticism: An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117–122.

Received April 22, 1957.

INTERPERSONAL PREFERENCE AND COGNITIVE ORGANIZATION[1]

NATHAN KOGAN AND RENATO TAGIURI

Harvard University

THE present paper describes an empirical test of certain aspects of Heider's theory of cognitive organization (2). In general terms, this theory states that the relation between interpersonal attitudes and social perception is largely determined by simple cognitive configurations. We shall deal exclusively with the case in which the elements of the cognitive unit are three persons—the perceiver, "p"; the other, "o"; and a third person, "q," and, in particular, with Heider's L-relations as represented by positive and negative feelings existing among members of small groups. These L-relations generate eight basic situations (cf. Table 1). Unit ※2, for example, corresponds to the case in which: p likes o, p dislikes q, and p thinks that o dislikes q.

According to Heider, units characterized by positive valences between the three component entities or by one positive and two negative valences are "balanced." Units containing three negative, or one negative and two positive valences are "imbalanced." With the assumption that people seek and attempt to maintain balanced cognitive structures while they aim to avoid and alter imbalanced ones, we shall test Heider's hypotheses by examining the incidence of two balanced and two imbalanced situations (Units ※1 and ※4, ※6 and ※7, Table 1). This necessarily requires the selection of appropriate baselines against which the observed frequencies can be compared. The baseline implicit in Heider's conceptual position is that all existing cognitive configurations will be balanced and none will be imbalanced. This is, however, an extremely harsh model for an empirical test. Furthermore, it fails to allow for the possibility of balance and imbalance as a continuous as opposed to a dichotomous process. There are two other baselines that provide relevant comparisons with observed values. The first, based on a chance expectancy model, is defined by the degree of balance that would be generated in the network if subjects were responding as random automata. The second baseline, based upon objective reality, is the actual level of balance present in the social unit. If cognitive balance exceeds this level, it would demonstrate that the cognitive process is sufficiently strong to impose a degree of balance that "overshoots" actuality. Our basic prediction is that *balanced* configurations will occur in *excess* of chance expectancy and objective reality. Correspondingly, we predict that the frequency of *imbalanced* configurations will be *below* these two baselines.

METHOD

Subjects and Procedure

Five groups of naval enlisted personnel with 12, 12, 15, 16, and 22 men, respectively, were used as *S*s. In each case, the men had been working and living together for at least two months on board ship as members of special teams.

The *S*s were requested, *first*, to indicate the three crew members with whom they would most like to spend a 72-hour liberty, and, *second*, to identify the three choices made by each of the other members. In the case of the 22-man group, which had been exposed to a great many research procedures, it was possible without unduly disturbing the members to obtain data regarding the three persons with whom they would *least* like to go on liberty.

RESULTS

Consider first the *balanced* Unit ※1 (cf. Table 1). If the perceiver, p, chooses two persons, o and q, then p may expect a total of two choices to exist between o and q, i.e., o's choice of q and q's choice of o. If, however, as is the case in our procedure, p chooses three other members of the group, o, q, and r, then p may attribute or expect that o choose q, q choose o, o choose r, r choose o, q choose r, and r choose q. That is, p may expect a total of six choices to exist among the three persons he chooses. If the data for all group members are combined, the total number of attributed choices expected would be six times the number of perceivers, that is, six times the number of members in the group (cf. "Total possible" column for balanced units in Table 2).

These maximum values represent perfect balance. As we suggested above, this is an exceedingly harsh baseline against which to

---

[1] This study was conducted at the Laboratory of Social Relations, Harvard University, under ONR Contract No. N5ori-07670. This paper was read at the EPA meetings, April 1957.

TABLE 1

COGNITIVE UNITS GENERATED BY HEIDER'S L-RELATIONS

Unit number	p's feeling for o (pLo)	p's feeling for q (pLq)	p's conception of o's feeling for q (oLq)	
1[a]	+	+	+	balanced
2	+	−	−	
3	−	+	−	
4[a]	−	−	+	
5[a]	+	+	−	imbalanced
6	+	−	+	
7	−	+	+	
8[a]	−	−	−	

[a] Units where pLo and pLq are of the same sign are transitive. Consequently, the unit includes both the oLq and qLo relationships.

TABLE 2

BASIC DATA FOR BALANCED AND IMBALANCED UNITS

Unit[a]	Group	N	Number of Choices: Total possible	Chance expectancy (A)	Perceived by p (B)	Actual (C)	Significance of differences based on chi-square tests, 1 df: A vs. B	A vs. C	B vs. C
Balanced #1	I	22	132	18.9	75	59	.001	.001	.01
	II	16	96	19.2	51	37	.001	.001	.01
	III	15	90	19.4	54	44	.001	.001	.05
	IV	12	72	19.8	42	25	.001	.20	.001
	V	12	72	19.8	33	26	.001	.20	.10
Balanced #4	I	22	132	18.9	47	34	.001	.001	.01
Imbalanced #6	I	22	198	28.3	16	19	.02	.10	.50
Imbalanced #7	I	22	198	28.3	7	6	.001	.001	.70

[a] Cf. Table 1 for description of specific units.

compare the observed number of attributions. Consider, instead, chance expectancy, one of the baselines mentioned above. If perceivers behaved in a random manner,[2] the number of choices p would expect to exist among the three members he chose would be $(3^2 \times 2)/(N - 1)$.[3] If the contribution of each perceiver is summed, the expected value would be $N(3^2 \times 2)/(N - 1)$. These expected values apply to both balanced units under consideration and are shown for each group in Column A, Table 2. A chi-square comparison of the total number of observed attributions with the number expected by chance in each of the five groups yields five ratios, all statistically significant beyond the .001 level (cf. A vs. B difference for balanced Unit #1 in Table 2). Clearly, the number of balanced configurations perceived by p deviates significantly from a chance baseline in the direction predicted by Heider's hypothesis.

Consider, next, the other baseline proposed above. Is the cognitive strain toward balance in excess of the corresponding relations existing in reality—i.e., the actual number of choices exchanged by the three chosen members? This baseline is quite stringent since the extent to which an *S*'s three favorites choose each other in fact already exceeds (significantly for the three larger groups) the chance baseline discussed above (cf. A vs. C difference for balanced Unit #1 in Table 2).[4] Comparison of attributed and actual choices shows, however, that cognitive balance exceeds objective balance in every case, and significantly so in four of the five groups (cf. B vs. C difference for balanced Unit #1 in Table 2).

Let us turn next to the *balanced* unit containing two negative and one positive valence (Unit #4, Table 1). According to the theory, if p *dislikes* o, q, and r, then p will tend to perceive these persons as *liking* each other. The relevant data for testing this hypothesis are available only for the 22-man group. The chance baseline here is identical to that described above for Unit #1. A comparison of the total number of attributions in the group with the expected chance value yields a chi-square significant well beyond the .001 level (cf. A vs. B difference for balanced Unit #2 in Table 2). Of greater interest, however, is the fact that the level of attribution again significantly exceeds actuality (cf. B vs. C difference for balanced Unit #2 in Table 2).

---

[2] The assumptions inherent in the present chance model are these. First, independence is assumed between the different selections made by any single individual. Second, the selections of any *S* are taken to be independent from those of any other *S*. Third, an *S* may not select the same other person more than once. Finally, he may not choose himself.

[3] The expected value is computed as follows. In a group of N members, o, q, and r must each choose three persons out of $N - 1$ others. The probability that o will choose q is then $3/(N - 1)$. Hence, the probability that both q and r will be chosen by o is $2 \times 3/(N - 1)$. The same set of operations is applied to q and r, with the result that the overall number of choices expected by chance between the members of any triad is $3(2 \times 3/N - 1)$ or $2 \times 3^2/(N - 1)$.

[4] This fact is in line with Newcomb's (5) suggestion that systems should show "symmetry" whether defined objectively or viewed through the eyes of a specific perceiver.

Finally, we turn to the two *imbalanced* situations (Units #6 and #7, Table 1), where data are available for the 22-man group only. According to Heider, if p likes o but dislikes q, then situations in which p perceives o as liking q should be extremely rare; correspondingly, if p dislikes o but likes q, p should very seldom perceive o as liking q. The analysis in the present case takes the form of counting the number of positive choices p perceives as directed by his three most liked to his three least liked companions, and the number of choices perceived as going in the opposite direction. Here the number expected by chance[5] for any perceiver is $3^3/(N - 1)$, or $3^3N/(N - 1)$ for the whole group (cf. Column A, Table 2, for imbalanced units). Chi squares for both imbalanced units (cf. A vs. B difference for imbalanced units in Table 2), show that imbalanced configurations occur significantly less often than would be expected by chance. The number of actual positive choices is also below chance expectancy, very significantly for #7 and at the .10 level for #6 (cf. A vs. C difference for imbalanced units in Table 2). The difference between perceived and actual values in these imbalanced cases yields negligible chi squares (cf. B vs. C difference for imbalanced units in Table 2). This last finding differs from that obtained for balanced situations, where, it will be recalled, attribution differed significantly from actuality. There is, however, a distinct "floor" effect in the imbalanced cases. While expected values are fairly low (28.3/198), actual values are even lower (19/198 and 6/198). Hence, there is very little "room" available for attributed values to fall significantly below the actual number of choices exchanged.

DISCUSSION

The present work applies the Heiderian model to the perception of affective feelings among members of small well-acquainted groups. The results support Heider's hypotheses regarding the cognition of affective relations. This evidence differs from, and thus supplements, the empirical studies of Heider's theory undertaken by Horowitz, Lyons, and Perlmutter (3) and by Jordan (4). In the former study, the emphasis is upon the evaluations of the *acts* of liked and disliked others. Jordan's experiment, on the other hand, while supplying data on affective relations, deals with *S*s' pleasantness-unpleasantness ratings of *imaginary* interpersonal situations verbally described.

We considered four of the eight cognitive units generated by all possible combinations of Heider's L-relations. In the uniformly positive balanced unit, perceivers attribute more affective links than expected by chance among the three members they prefer most. More important, however, is the fact that the attributed number of choices is also significantly in excess of the actual number existing between chosen others. Hence, the cognitive disposition toward balanced units, while deviating from chance in the direction of veridicality, is sufficiently strong to be a source of some systematic perceptual error. This "overshooting of actuality" was also observed in the other balanced situation under investigation, where the perceiver assumes that those he dislikes tend to choose one another more than they actually do.

The two imbalanced units examined also conform to Heider's prediction. In both cases, these cognitive configurations have frequencies significantly below chance, pointing to the rarity of imbalanced cognitive structures. It should be noted, however, that this veritable absence of cognitive imbalance is in line with actuality. Apparently, this is a very difficult state to maintain in the actual social microsystem as well as in the cognitive representation of this system. While the occurrence of imbalanced configurations is rare, the differential frequency of the two types of imbalance is worth noting. It appears to be cognitively more difficult to conceive that a person we like will choose one we do not like than it is to conceive that a person we do not like will choose one we like. The present evidence implies that a simple balance-imbalance dichotomy does not do complete justice to the data. Rather, as Cartwright and Harary (1) also point out, there are degrees of balance and imbalance. Given the possibility, with the present methodology, of distinguishing degrees of cognitive and actual balance and imbalance, individual differences in cognitive dispositions toward balance become accessible to research.

---

[5] The expected value is computed as follows. Consider a group of N members and randomly select any two triads whose members are o, q, r and o′, q′, r′. Each must choose three persons out of $N - 1$ others in the group. The probability that o, for example, will choose q′ is $3/(N - 1)$, and, hence, the probability of o choosing o′, q′, and r′ becomes $3 \times 3/(N - 1)$. Given identical probabilities for q and r choosing o′, q′, and r′, the overall number of choices expected by chance directed from the three members chosen by p to the three members rejected by p is $3 \times 3 \times 3/(N - 1)$ or $3^3/(N - 1)$. This expected value is obviously also applicable to the other imbalanced case in which choices are directed from a rejected triad of members to members in a chosen triad.

An adequate conceptualization of the problem of cognitive balance must deal with the degree of correspondence between cognitive systems and systems objectively defined, a matter relatively ignored in discussions of Heider's theory. Is it legitimate to speak of a strain toward cognitive balance or away from cognitive imbalance in those cases where the cognitive and objective systems are highly similar? Could we not explain the situation more simply in terms of the *S*s' *accurate* awareness of reality? The shortcoming of an accuracy explanation in the present context is that it fails to account for the fact that strain toward cognitive balance may manifest itself in the very choice (or rejection) behavior of a member. We have shown that the level of affective compatibility among members chosen or rejected by a person is already higher than that of a randomly assembled triad. This very fact implies that cognitive determinants are involved and supports the Heiderian principle of preference for balanced and aversion for imbalanced units.

Heider's theoretical system appears to be a most useful vehicle for exploring the problem of interpersonal feelings and perceptions. The present paper demonstrates that the hypotheses derived from the theory are confirmed when empirically tested in a small group situation and that new research problems emerge from the application of Heider's conceptions to this type of data.

SUMMARY

The perception of interpersonal preferences in small groups represents a problem area of immediate relevance to Heider's theory of cognitive configurations. In the present investigation, hypotheses derived from this theory are empirically tested and confirmed. Balanced cognitive units occur significantly in excess of chance and actuality baselines. Imbalanced cognitive units, on the other hand, appear significantly less often than would be expected by chance, though it should be noted that actual social units of this kind are also very rare. Theoretical and experimental implications of the findings are discussed.

REFERENCES

1. CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63**, 277–293.
2. HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.
3. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57–76.
4. JORDAN, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6**, 273–287.
5. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.

Received April 22, 1957.

CASE REPORT

SOCIAL BEHAVIOR OF YOUNG CHILDREN IN THE *KIBBUTZ*[1]

HELEN FAIGIN

The Hebrew University, Jerusalem

THE system of collective education followed in the *kibbutzim* (collective settlements)[2] in Israel provides a unique opportunity for the study of the effects on personality development and social behavior of a pattern of socialization that stresses group living and the importance of peer relationships from the very beginnings of life. The study reported here is an exploratory investigation of the social relationships among children of toddler age growing up in this setting.

Within the system of collective education, the children are reared in groups from birth through adolescence. They live in their own houses apart from their parents and are cared for by a *metapelet* (children's caretaker) whose responsibility it is to care for and socialize the children. The role of the *metapelet* (plural, *metaplot*) changes with the age of the children and as the children grow older, teachers and the children themselves assume some of the roles which the *metapelet* had performed earlier. The children spend two to three hours every afternoon with their parents and most of the day on Saturday, the day of rest. Whereas the *metapelet* (especially in the early years) has responsibility and authority for the socialization of the children, the role of the parents is to be "good friends" with the children. The parents' role as disciplinary agents is negligible.

Toddlers' groups were chosen as the special focus of the study in order to determine whether under these special conditions of group-living social interaction between children and group identification can be learned in very young children. Most studies and observations of children in western societies indicate that children interact with one another to a very small degree and for short lengths of time until approximately the age of three years. There is little concept of group or group identification at a younger age. The children are oriented for the most part to themselves and to the significant adults in their environment. The development from a biological to a social being is a slow and gradual process. Do children living in groups from birth in this type of setting learn to interact with each other at an earlier age; do they begin to feel themselves as part of a group and develop close relationships with one another earlier than children reared in families? These were the primary questions investigated in this study. A secondary point of emphasis was the role of the *metapelet* and her relationship with the children.

METHOD

The Sample

The study was carried out in two settlements, in each of which three groups of toddlers were observed. The children ranged in age from 19 to 38 months at the time the study was begun. The groups were divided according to age, with an age span of five to six months within each group. All groups consisted of six children.

The two settlements had been in existence for approximately twenty years at the time of the study. The members of both *kibbutzim* are of European origin and have the same general educational level. Each of the *kibbutzim* belongs to a different *kibbutz* movement[3] and can be said to be typical to a certain degree of the educational philosophies and practices of that particular movement. The differences in educational orientation may be summarized in terms of the greater degree of deliberate educational stimulation and the more structured setting of *Kibbutz* S, as compared with the greater freedom allowed the child in *Kibbutz* M.

---

[1] This research was supported by the Committee for Research on Collective Education of Seminar Hakibbutzim.

[2] For description of the *kibbutz* setting and the system of Collective Education, see (2, 6, 7).

[3] Each *kibbutz* is a member of a *kibbutz* movement which coincides in general with a political party. Each movement has certain political, social, and educational ideologies which distinguishes it from the other movements. For fuller detail see Spiro (7).

Method of Observation

Observations were made on the children at regular intervals over a period of six months. Systematic observations were made on the aggressive and dependent behavior of each individual child[4] in various activities: free play, organized play, meal time, and getting up in the afternoon. In all, there were three hours of systematic observation made on each child in ten- or fifteen-minute time units. Specific time periods were allotted to each activity. Observations were made, also, upon each group as a whole in the above noted activities and, in addition, getting up in the morning, going to sleep in the afternoon and evening, on walks, at shower time, etc. In this way, the children were observed in all the daily routines. In addition, both formal and informal discussions were held with the *metaplot*, parents, and the person or persons responsible for organizing and directing educational practices with young children in each of the settlements.

Nature of the Data

Systematic observations of dependent and aggressive behavior of the children were made in terms of predetermined categories to be described, so as to permit statistical analysis of the frequency of different response categories.

Most of the observations were made by the writer. Help was obtained for a short period of time from two assistants in order to establish reliability of observation.[5] The reliability of observation of the various categories of aggressive responses ranged from 70 to 95 per cent, using the formula:

$$\text{per cent agreement} = \frac{\text{number of agreements of two observers}}{\text{number of agreements plus number of disagreements}}$$

Agreement on total aggression was 92 per cent. The reliability of the dependency categories ranged from 75 to 97 per cent. Agreement on total dependency was 90 per cent.

The data on group behavior were obtained from consistent rather than systematic observations. Full records were kept of the occurrences of responses of the various categories of social behavior observed. Special care was taken not to generalize from one instance of unusual or interesting behavior but to generalize only from types of behavior that occurred with some frequency over the entire period of observation. These data were used for descriptive purposes. All data on the behavior of the children are based on these two types of observation. Information was thus obtained on the day-to-day activities of the children, the patterns of interaction among them, and their interaction with the *metapelet*.

The data obtained from formal and informal discussions with *metaplot* and educators were used as background material and as a basis for assessing their attitudes toward the educational system and the problems, general and specific, that arise in taking care of very young children.

THE DAILY ROUTINE

The daily routine in the toddlers' house is quite similar in the two *kibbutzim*, but the daily program is somewhat more set and routinized in *Kibbutz* S than in *Kibbutz* M. The children wake up in the morning at about 6:00 or 6:30. There is a period of free play before breakfast usually inside the house. After breakfast there is again free play; the children usually play outside while the *metapelet* cleans up the house. After this the children usually go on walks around the settlement, visiting various work branches, go swimming in the summer, have their midmorning snack, play again around the house, and then have lunch. In *Kibbutz* S, almost every morning after the walk, or in place of it, the children spend about half an hour in concentrated activity known as "work": in general, some creative or constructive activity, such as drawing, building, or reading. Emphasis is placed on working well and quietly. This morning "work" period tends to become a routinized daily activity when the children are about 24–27 months old. In *Kibbutz* M, concentrated activity is an almost daily occurrence in the oldest (3-year) group and occurs once in a while in the younger groups. The concept of "work" is not employed and there is not as much emphasis on quiet concentrated activity.

---

[4] There were six children in each group, but the individual child sample included only those children born in the *kibbutz*. This made a total of 31 children: five children from five of the groups, and six children in one of the groups.

[5] These assistants were seminary students, who were given special training in the observation and recording of aggression and dependency, including ten hours each of observing dependency and aggression in groups of children of similar age but not in the research sample, during free play activity. The data on reliability summarized here were obtained from ten hours of joint observation by the author and an assistant. Of the five hours that each assistant observed with the author, three hours of observation were made during the free play period. Observations were divided quite equally among the three groups of children.

The children eat lunch at about 11:00 A.M. and then go out and play in the yard while the *metapelet* cleans up. After this they shower and take their afternoon nap, which usually lasts around two hours. The children are washed, dressed, and given a snack, and then go off to their parents. In *Kibbutz* M, the children have their evening meal at this time. The reason for this will be explained below.

The children spend two to three hours with the parents and siblings. The parents (one or both) bring the children back to the children's house. In *Kibbutz* S, the children have their evening meal, and the *metapelet* puts them to bed. In *Kibbutz* M, the parents bring the children back to the children's house and put them to sleep. In both *kibbutzim*, the children sleep in the houses all night without any adult with them. A night watchwoman, who is assigned to the children's houses, walks about in the area and looks after any children that cry or have any particular problems.

THE ROLE OF THE *METAPELET*

The role of the *metapelet* of the toddlers includes a large number of duties and responsibilities; the *metapelet* functions as housekeeper, caretaker, socializing agent, and teacher. She does all the things that a mother usually does and has been delegated the responsibility and authority for rearing the children. Her evaluation as a *metapelet* is in terms of how well the children in her care develop. A close attachment exists between the children, and the *metapelet* and the children will often go to visit the *metapelet* who worked with them at this age when they are older and she no longer works with them.

The Metapelet as Housekeeper

One of the *metapelet's* duties is to keep the children's house clean and in order. She washes the floors, the dishes, makes the beds, takes care of the laundry and in general keeps the house in order. To this task the *metapelet* devotes more or less time depending upon the age of the children, upon her own interests, her efficiency as a housekeeper, and how she herself defines her role.

The Metapelet as Caretaker

Physical care of the children. The *metapelet* feeds, washes, bathes, and clothes the children. She takes care of them when they are ill, calls the doctor if necessary, does all that is necessary for their general health.

Emotional care of the children. This is the area around which many of the problems of collective education are centered. There is much discussion, both by parents and by *metaplot*, about whether or not the children get enough individual attention. Since the group is composed of six children, the problem of balancing individual needs against group needs often arises. There are special courses for *metaplot*, and many discussions are held about the emotional needs of the young child in an effort to help with the problem arising in this area. The *metapelet* usually tries to give the child as much attention as he demands; the children who demand more usually receive more. The *metapelet* often spontaneously hugs and kisses the children when she is bathing and dressing them, holds their hands when they go for a walk, and kisses them when she puts them to sleep. There are individual differences among the *metaplot* in "warmth" and "motherliness" which influence the amount of attention and affection that they express toward the children.

There exists some conflict about the *metapelet's* infringing on the parents' loving role. The *metapelet* often rationalizes her feelings of not giving the children enough individual attention by saying that they get enough from their parents. On the other hand, the parents tend to resent the child's strong attachment to the *metapelet*.

The Metapelet as Socializing Agent

Among the important duties of the *metapelet* is the socialization of the basic disciplines. Weaning is done by the mother, but after this the responsibility is largely the *metapelet's*. She teaches the children to eat by themselves, toilet trains them, teaches them to wash and dress themselves, answers their questions about sex, permits and limits their expression of aggression, and in general makes demands upon them in all areas of behavior.

Eating. The children begin eating by themselves at about one year of age. At this time, little fuss is made about the child smearing the table or his face. There are very few feeding problems. Since the children tend to imitate one another, learning to eat in a group seems to make the process easy and quick.

Toilet training. Toilet training is begun when the child is about eighteen months old. The children sit down in a group together, each on his own pot. The company of the other children and the presence of the *metapelet* seem to make toileting one of the activities of the daily routine. In each of the youngest groups (under two years of age) there were two children who did not always make a bowel movement in the pot. All the children in the other groups were bowel trained. Most of the children, even in the youngest groups, usually urinated in the pot during the day; occasional accidents occurred when the child was involved in his play. Two children in the oldest groups (two-and-a-half to three years) occasionally wet at night, the others were all dry. In the middle groups (two to two-and-a-half years) most of the children occasionally wet at night, but none wet every night. In the youngest groups, (under two years) there were four children who wet every night, the rest, occasionally.

Sex training. There is no separation of the sexes in the groups. The children sleep, bathe, and go to the toilet together. The children masturbate sometimes when sitting on the pot, when standing about doing nothing, and occasionally when upset. The children masturbate each other occasionally. The *metapelet* usually pays no attention to the child's masturbation, but occasionally tries to distract him by getting him interested in some activity.

Independence training. The children begin to try to dress themselves by about the age of two years. There is no direct pressure for them to do so, but they are highly praised for their attempts in this direction. They eat alone at an early age, and in general are encouraged to do things by themselves. They are praised for their achievements. Emphasis is placed on "being big." There is somewhat more emphasis on independence in *Kibbutz* S than in *Kibbutz* M.

Control of aggression. The problem of aggression is discussed later in detail. In general, the *metapelet* does not interfere in quarrels between the children, only in cases when a child is getting badly hurt. The general attitude is that the children should have the opportunity for working through their quarrels by themselves.

Discipline. Physical punishment is taboo in the *kibbutz*. Discipline can be characterized, in general, in terms of explaining and trying to win the child over by suggestion or persuasion. A very effective technique is to separate the child from the other members of the group until he is ready to behave properly. Other members of the group often join in and suggest that the "erring" child be deprived of something or they tell him what to do.

The children are very cooperative in general with regard to the demands made upon them, and there are no real disciplinary problems in these groups. Disobedience and disciplinary problems arise much more frequently when the child is with his parents. The general atmosphere within the groups is permissive, and few demands for obedience are made upon the child. There is a tendency to conform to group demands and pressures.

The Metapelet as Educator

Education as used here refers to the active teaching and stimulating of the children as opposed to the subtle, barely conscious teaching of attitudes and relationships. The *metapelet's* role as educator is greatly stressed in *Kibbutz* S, whereas in *Kibbutz* M it is hardly considered a part of her role with these young children.

The *metapelet* in *Kibbutz* S also sees guiding and educating the parents in how to handle their children as an important part of her task.

THE CHILDREN'S GROUPS

The children's groups are described in terms of the general tendency toward group identification, the quality and types of interaction, and the effects of group living on the social development of the children.

Identification with the Group

There seems to exist a certain strong "we group" feeling in all of the children's groups. This was evident in their supporting each other whenever criticized or attacked by children from another group. All toys, furniture, and space within the children's house were referred to as "ours" as opposed to "theirs." There were many instances of out-group aggression, and the group united to support its own members. The children often did not permit children from other groups to play with their toys or to come into their house. Whenever a child was absent from a group activity, his absence was noted and mentioned. For example, when a group goes for a walk, even the youngest groups always see to it that all the children are present before going off, without any prompting from the *metapelet*. One of the favorite games of the children was playing on a carousel. They always made room for the members of their own group, but would not permit children from other groups to join them.

The concept of "we" and "ours" was well developed in the children by the age of two. All activities and property connected with the children's house were referred to by the children as "ours," whereas anything that a child had or did in his parents' room was referred to as "mine." A child might bring candy from his parents' room to the children's house and it was up to him whether or not he would distribute it; if he did, he usually did so with much ceremony. The general idea is that this is mine and I am sharing it with you. He does not have to share, however, and no pressure is put upon him to do so by the *metapelet*. The attitude toward things around the children's house, however, was quite different. Here, the expected thing was to share—everything is *ours*. Even when a child had a coveted toy which he did not want to give to another child—he would not say "this is mine"—he would say, "but I had it first."

The children in one group would often brag about the things they had done or the property they had in their house to children of another group. "We have a nice house and you don't," "we went for a walk and you didn't," and the children of the other group would retort in kind.

If a child was hit or teased by a child from another group, the other children in his group would always rise to his defense no matter what the reason was for the quarrel. The children in a given group would often adopt the same attitudes toward things: such and such food is good, such and such is not, such and such toy is good, such and such is not, such and such person is good, such and such is not.

When a child did not want to do something—for example, eat, help clean up, look at a book, in general to do the same activity as the rest of the group—the *metapelet* would usually say, "If you don't want to you don't have to, you can go out and play." This was reacted to by the children as punishment and they joined in the activity.

The children often imitated one another at mealtime and in many of the group activities. Often when the *metapelet* helped a child eat or dress, this child would tell her to help the other children also.

In *Kibbutz* S, this group orientation and identification was more verbalized both by the *metaplot* and by the children. The *metapelet* very consciously talked about the group, in many ways encouraged the development of the concept "we" and "ours." In *Kibbutz* M, this was done much less consciously, but the group feeling exists. A favorite game of the children in the oldest group in *Kibbutz* M, when the observer was present, was to ask her to write down their names. Each child would ask to have his name and the names of all the other children in the group written down. No child ever omitted the name of any of the other children in the group. In both settlements, the children always made sure that everyone was present at mealtime, and at all group activities, such as going for walks and concentrated activity such as reading, drawing, and making things out of plastalina. This making sure that everyone is present was more frequent and consistent in *Kibbutz* S.

The more conscious awareness and expression of group feeling in *Kibbutz* S than in *Kibbutz* M seems attributable mainly to two factors: (*a*) the greater general awareness and articulation in the *Kibbutz* S movement of inculcating the ideology of group living in the young children and (*b*) in *Kibbutz* M, the groups break up at the age of three to three-and-a-half to go to separate kindergartens and come together again only in the second grade—whereas in *Kibbutz* S the group continues to stay together.

Social Participation in the Children's Groups

Much of the time the children spend playing separately, and yet together. Each child is involved in his own particular activity, but plays in the vicinity of the other children. Each child looked around from time to time to see if the others were present and would often point out to the other what he was doing. If any child was not present, the other children would inquire as to this child's whereabouts.

Even in the youngest groups the children often played games together such as peek-a-boo in a group or just two or three of them together. They often imitated one another, crawled on the floor, or pretended that they were dogs. Often two or three children would evolve some game which might go on as long as 15 minutes. In the youngest group, when they played in twos and threes, the specific children changed from activity to activity and from game to game. In the two older groups, certain friendships had formed and there were often two or three children who usually played together. There began to be a trend toward pairing off or forming smaller groups within the group. This was observed mostly during free play.

In the youngest groups, "play together"—aside from general group activities such as going for walks or riding on the carousel—usually involved looking at things such as grass or insects together—and laughing, playing peek-a-boo, imitating animals, and occasionally building things together. In the middle groups, the children often build things together and play more involved games such as picking and exchanging flowers, exchanging toys, and building canals together out of sand and water. In the oldest group, the children built things together, and there was also quite a bit of imaginative play in which the children took on roles of adults. The most popular of these games in the oldest groups in both *kibbutzim* was playing mother, father, and baby. The same game would be continued by the same group of children day after day, during some of the free play time. Usually, the same child would play the same role each day; sometimes there would be shifts in the child who played a particular role. It is interesting to note, however, that this imaginative play centered on the family, not on the children's group and the *metapelet*. Occasionally, the children played doctor and similar games, but the family group was most popular. Further investigation of this problem in terms of analysis of the children's play might prove rewarding in getting at the kinds of relationships these children have with their families as compared to children growing up in private families.

Group Values

One of the important values expressed by the children both in their behavior and verbally in all the groups in both settlements is sharing and taking turns. This was apparent especially with regard to food, and playthings such as paper, crayons, and plastelina. The children knew that they would each get some and that they had to wait their turn. How much the children learned in this area was especially evident when a new child came into the group who had not learned to share and wait his turn. The other children explained to him and often pounced on him when he did not behave properly.

In *Kibbutz* S there was also much stress put upon being big and strong and independent, and on doing "constructive" things—"working" quietly and making things. The "big" and independent child can take care of himself, can dress himself, and doesn't cry when he gets hurt. In *Kibbutz* M, children were praised for these things, but there was little stress placed on them.

Values expressed with regard to fighting and displaying affection are discussed later.

The children's groups, even the youngest, play an important part in setting standards for behavior and in controlling the behavior of group members. Often the children's statements of encouragement or discipline were direct imitations of the words of the *metapelet*; sometimes they were inventions of the children themselves. Whichever they were, they were used at the appropriate times and carried the weight of sanctions. The children to whom the statements were addressed responded in accordance with them.

In the youngest groups in both settlements, most statements of social control referred to eating and toilet training. Some times they also referred to matters of disobedience and aggression. For example, when one of the children made a bowel movement on the floor the other children in the group looked at him and they all said, "Haggai did a BM, Haggai did a BM, not on the floor, Haggai, not on the floor." At another time the children were sitting around the table eating soup. One of the children said the soup was hot, so another child said to her "It's not hot, taste it and you'll see, it's not hot." At times, this control took the form of action rather than words. One of the children had been hitting another child; the second child was crying. All the rest of the children in the group came over and hit the first child.

In the older groups in both *kibbutzim* this control covers all areas of behavior. Sometimes the older children told the younger children how to behave. Below are several examples of statements used by the children in the two older groups to control each other's behavior. First, statements referring to eating:

One of the children isn't eating. Another child says, "Why aren't you eating—I'll feed you." He held him by the arm and fed him.

When one of the children did not want to eat her soup, another child said to her, "It will cool off soon, then you should eat it, when it cools off."

One of the children wanted the observer to write his name, but he hadn't yet eaten his lunch. One of the other children said, "She won't write it for you until you eat." All the other children joined in and told him that first he had to eat and then she (the observer) would write his name.

One of the children wanted to play with nails. Two of the other children told the *metapelet*, "Don't give him any nails yet, he hasn't finished eating."

Statements referring to aggression:

Two children were standing in the house looking out the window. They saw another child hitting a baby. One said to the other, "One mustn't hit babies." Then he called out to the child outside, "You mustn't hit the baby."

One child said to another, "I'm going to hit you." The second child said, "Why do you want to hit me?" The first child then said, "I won't hit you" and hit the table instead.

One child hit another. A third child came over and hugged the child who had been hit and asked, "Who hit you?" He then told that child that he must not hit the other child again.

One child hit another. The second child said, "Why are you hitting me? I'm not hitting you." The first child stopped hitting.

One child hit another. A third child told the child who had hit, "How many times do I have to tell you not to hit other children."

Social Control

Statements referring to being careful:

One of the children put a pin in his mouth. Another child told him, "You mustn't put a pin in your mouth—you'll get a sore." She took the pin away from him.

On sharing:

Two children were fighting over who was to sit by the steering wheel in an auto. A third child said: "Giladi, that belongs to all of you, to Yisrael too, he can sit there, too, both of you can." They both sat down beside the steering wheel.

The children from a very early age have thus taken over the values of the culture in which they are being reared and use these actively in setting standards for each other's behavior. Not only are these social control statements used but they carry weight in controlling behavior of the children. So the group operates as a socializing agent upon its members from a very young age. In many instances the forms of social control expressed and exercised by these young children seemed to have the quality of "punishment by reciprocity" and to carry the notions of "equality" and "solidarity" as these concepts are used by Piaget (4). These children are much younger than the age at which the idea of distributive or equalitarian justice are thought to develop, but perhaps the conditions of group living tend to induce the development of cooperation, equality, and group solidarity at a very early age. A more focussed study of the methods of social control used by very young children would seem highly desirable.

It is interesting to compare these observations with the use of rules by three-year-old middle-class nursery school children studied by the same investigator (1) in a small city in the United States. Two interesting differences may be noted. The American three-year-olds tended to use more rules about property, that is, about toys and play space, than rules of interaction, how to behave. In the *kibbutz* group, the children although younger tend to use more rules about social interaction than property. This difference would seem to follow from the fact that the *kibbutz* children live and interact with one another to a much greater extent than do nursery school children. The other interesting difference concerns the nature of the property rules. The American nursery school children used rules referring to this being "mine"—"I had it first, you can't have it"—whereas the *kibbutz* children's property rules refer to sharing, "This is ours, this belongs to all of us." The social control methods used by these very young children thus appear to reflect the values of the society in which they grow up: one a society in which private property is highly valued, the other in which there is communal ownership and no private ownership.

Relations Between the Children

The affective ties between the children are very strong. They do not approach the intensity of relations observed among siblings outside the *kibbutz*, and yet they are much more intense than that of friendships among other Israeli children of this age who play together. In spite of the fact that much of the time the children play by themselves, they almost always play in the vicinity of the other children and are carefully attuned to the movements of the other children. They are aware of the presence and absence of the other children, know what they are doing, respond immediately to the crying and fighting of the other children.

Many of the *metaplot* and parents have some doubts about whether or not children of this age should be living in groups. They talk about the "struggle for existence" which goes on among the children. They say that there is too much fighting and that the strong, more aggressive children hold the power and can get what they want, whereas the weaker, less aggressive children have a hard time holding their own and have to give in to the stronger. This is only partially the case. There is much fighting and crying and the stronger children often do hit the weaker children and take the toys away from them. On the other hand, the children often comfort each other when they are upset or crying. When a child is crying because he has been hurt or because a toy was taken away from him, very often the aggressor will come over and pat the child, tell him he will not hurt him again and perhaps return the toy he has taken or give him another in exchange. If the child who has been the aggressor does not comfort the hurt child, other children in the group will do so.

In each of the two youngest groups there was one extremely aggressive child. These children very often hit the other children and took toys from them with seemingly no provocation, at least no provocation apparent in the situation. The other children tended to avoid them, and from time to time would gang up on the aggressive child. In the two older groups in each of the *kibbutzim* there was no child who could be called extremely aggressive, in the sense of showing behavior similar to the two children in the youngest groups.

In none of the groups was there a domineering or bossy child who would try to tell all the others what to do. During the course of the observation, a new child came into one of the groups who tried to boss the other children around. She quickly learned that she could not do this. The other children either paid no attention to her or fell upon her as a group and beat her. In a very short time this behavior disappeared.

There was no child in any of the groups who could be called the leader. At different times different children played leadership roles in given activities, but no child played this role consistently. Leadership was always in the hands of the *metapelet*. Often when older children would come to visit, an older sibling of one of the children in the group, this older child would take a leader role, organize some activity and the children in the group would follow this leadership.

SYSTEMATIC OBSERVATIONS ON AGGRESSION AND DEPENDENCY

Aggression

It is difficult to give a general picture of the frequency or types of aggression expressed by the children in the two settlements. Systematic observations were made of the frequency of various types of aggression expressed by the children during free play. Each child was observed at four different periods of 15 minutes each. Nothing can be said about whether children in *kibbutzim* display more or less aggression than children growing up in private families, since there are no comparative data. The data have been analyzed to compare the children in the two *kibbutzim*, with special emphasis on the different educational practices, and to compare the different age groups within each *kibbutz*.

TABLE 1

MEAN AGGRESSIVE RESPONSES FOR THREE GROUPS OF CHILDREN IN EACH *Kibbutz*

	Mean Number of Responses per Child During Four 15-Minute Periods of Observation							
	Kibbutz M				*Kibbutz* S			
	Grp. A[a]	Grp. B	Grp. C	All Grps.	Grp. A	Grp. B	Grp. C	All Grps.
Category	(N = 5)	(N = 5)	(N = 5)	(N = 15)	(N = 6)	(N = 5)	(N = 5)	(N = 16)
Physical aggression	4.2	3.4	1.2	2.9	2.5	2.4	2.6	2.5
Instrumental aggression	1.0	3.4	1.8	2.1	4.3	1.0	1.0	2.2
Defense of toys	0.2	0.8	1.4	0.8	0.5	0.8	3.0	1.4
Defense of self	0.2	0.8	0.6	0.5	0.2	1.2	1.2	0.8
Aggression to objects	1.2	2.0	0.0	1.0	2.8	2.0	3.8	2.5
Verbal aggression	0.0	0.0	2.0	0.7	0.0	0.2	0.8	0.3
Aggression to smaller children	0.0	0.8	2.2	1.0	0.0	0.4	2.8	1.0
Disobedience	0.4	1.0	0.4	0.6	0.3	0.4	1.8	0.8
Total aggression	7.2	12.2	9.6	9.7	10.7	8.4	17.0	11.9
(*SD*)	(6.2)	(6.8)	(7.2)	(6.6)	(11.2)	(5.8)	(11.5)	(10.6)

[a] In each case, Group A is composed of the youngest children, Group C of the oldest.

The aggression shown by the children was broken down into various types of aggression. The data on the frequency of all types of aggression recorded are shown in Table 1.

Physical aggression. Physical aggression is defined in terms of hitting for the purpose of hitting or hurting. There were no statistically significant differences between the children in the two *kibbutzim* in the frequency of use of physical aggression as can be seen from Table 1. Within the *kibbutzim* the only significant difference between age groups was found in *Kibbutz* M in which the children in the oldest group exhibited significantly less physical aggression than the children in the middle age group. The highest frequency of expression of physical aggression was by a child in the youngest group in *Kibbutz* M, one of the very aggressive children previously described, for whom 15 units of physical aggression were recorded in the hour's observation. For many of the children, 6 out of 31, no physical aggression was recorded during the hour period, and for nine children only one instance of physical aggression. Thus, about half the children (15 out of 31) expressed very little or no physical aggression.

Instrumental aggression. Instrumental aggression is defined in terms of hitting or pushing in order to attain some other goal, such as a toy, play space, or attention. Table 1 shows the frequency of expression of instrumental aggression. There were no significant differences between the *kibbutzim* or between the various age groups. The highest frequency of instrumental aggression was recorded for one of the previously described aggressive children, in the youngest group in *Kibbutz* S, 20 units. For 13 children, no instances of instrumental aggression were recorded, and for five children only one instance of instrumental aggression was recorded. Thus, for 18 out of 31 children there was practically no instrumental aggression expressed.

Other forms of aggression. The frequency of occurrence of the other categories of aggression (defense of toys, defense of self, aggression toward objects, verbal aggression, aggression toward younger children, and disobedience) is so small that not much can be said about it except to note its lack of occurrence.

Total aggression. Total aggression includes all forms of aggression, the total scores for the various types of aggressive responses. There were no significant differences in the frequency scores of total aggression (Table 1) between the two *kibbutzim*, between boys and girls, nor among the different groups. There were large individual differences in the frequency of occurrence of aggressive responses as reflected in the large *SD*'s, presented in Table 1, with a range from 1 to 39 in the frequency of total aggression.

These wide individual variations would seem to indicate that the instigation to aggression is due to factors other than or in addition to the frustrations inherent in group living. It was observed that not only did the individual *metaplot* respond differently in general to the aggressive responses of the children and by their patterns of relationship instigate more or less aggression, but each *metapelet* responded differently to each individual child when he responded aggresively. Some children were encouraged to respond aggressively, some discouraged from this, and some "just got away with it." Although the *metaplot* tend to practice a "hands off" policy, they were frequently inconsistent in their reactions to the aggressive responses of individual children. The presence or absence of the *metapelet* at the time of occurrence of aggressive responses adds to the factor of inconsistency. In addition, the children themselves respond differentially to one another in reaction to aggression directed toward them. The family constellation, too, probably has an important effect on the child's tendency to respond aggressively. It was noted, however, that children who are less aggressive in the group are more aggressive in the parents' room and vice versa.

Dependency

The dependency behavior expressed by the children was observed and recorded systematically as was aggressive behavior. Under the general heading of dependency the following types of behavior were recorded: crying, showing affection toward adults (other than the *metapelet* and observer), showing affection toward the observer, showing affection to the *metapelet*, showing affection toward children, instrumental dependency, i.e., asking help to do something, seeking attention or approval for making or doing something, and thumb-sucking. The frequencies with which these responses occurred in the various groups in the two *kibbutzim* may be seen in Table 2.

Expressing affection was categorized under dependency since it is also a means of seeking attention and affection. Affection is expressed by responses such as touching, holding, or being near. The nature of these responses inevitably induced a response of a similar type from the other person—thus the child's giving and seeking attention could not be separated.

Dependency toward adults. The one consistent difference that appeared in all the categories of dependent behavior (except affection toward children and displaying achievements) was that between the children of the youngest groups in the two *kibbutzim*. In total dependency toward adults, the children in the youngest group in *Kibbutz* S showed higher scores than the children in the youngest group in *Kibbutz* M.[6] This same type of difference was found in the separate categories of affection toward adults, toward *metapelet*, toward observer (.10 level of confidence) and instrumental dependency (.10 level of confidence). In contrast, the children in the youngest group in *Kibbutz* M cried more frequently (.10 level of confidence) and sucked their thumbs more frequently (.10 level of confidence) than did the children in the youngest group in *Kibbutz* S.

These apparently inconsistent findings probably can best be understood in terms of the differences in educational orientation in the two *kibbutzim*. It will be remembered that *Kibbutz* S was described as stressing educational stimulation more at an early age, whereas *Kibbutz* M tends to stress a great deal of freedom and permissiveness, letting the child develop at his own rate. One of the consequences of this difference is that the *metapelet* in *Kibbutz* S spends more time with the children during free play and hence they direct their dependency responses toward her. In contrast, the children in *Kibbutz* M tend to cry more unheeded and suck their thumbs since the *metapelet* is not around. There is less opportunity to direct dependency responses toward adults. In general, the children in the youngest group in *Kibbutz* S seemed to develop at a faster rate than the children in *Kibbutz* M. They seemed to be more aware of adults, more aware of their surroundings, played more, and spoke earlier than the children in *Kibbutz* M. These differences in rate of development seem to level off in the two older groups where the same differences do not occur.

Differences were found between the oldest groups in the two *kibbutzim* in two categories: displaying achievements and instrumental dependency. Here the children of *Kibbutz* M more frequently displayed this behavior than the children of *Kibbutz* S, contrary to what might have been expected. A possible explanation is that the *metapelet* of the oldest group

---

[6] All differences mentioned are statistically significant at the .05 level of confidence using the t test, unless otherwise stated.

TABLE 2

MEAN DEPENDENCY RESPONSES FOR THREE GROUPS OF CHILDREN FROM EACH *Kibbutz*

	Mean Numbers of Responses per Child During Four 15-Minute Periods of Observation							
	Kibbutz M				*Kibbutz* S			
	Grp. A[a]	Grp. B	Grp. C	All Grps.	Grp. A	Grp. B	Grp. C	All Grps.
Category	(N = 5)	(N = 5)	(N = 5)	(N = 15)	(N = 6)	(N = 5)	(N = 5)	(N = 16)
Crying	10.6	3.4	2.8	5.6	4.7	4.0	2.4	3.8
Affection to adults	1.8	7.2	6.8	5.3	5.8	9.0	8.0	7.5
Affection to observer	8.0	9.0	14.2	10.4	21.3	8.8	6.4	12.8
Affection to *metapelet*	0.8	4.2	6.4	3.8	2.8	9.4	8.4	6.6
Instrumental dependency	0.4	0.6	5.8	2.3	1.3	1.6	1.2	1.3
Displaying achievements	0.0	0.6	2.8	1.1	0.0	2.6	0.6	1.0
Total dependency to adults	21.6	25.0	39.0	28.5	35.8	35.4	27.0	32.9
Affection to children	1.0	5.6	4.0	3.5	1.3	2.0	4.0	2.4
Thumbsucking[b]	7.2	6.2	2.0	5.1	1.5	6.0	1.2	2.8
Total dependency	22.6	30.6	43.0	32.1	37.3	37.5	31.0	35.3
(*SD*)	(10.4)	(5.0)	(17.4)	(15.2)	(9.5)	(20.6)	(19.0)	(16.9)

[a] In each case, Group A is composed of the youngest children, Group C of the oldest.
[b] Not included in total dependency score.

in *Kibbutz* S usually praised the achievements of the children without their asking for it. The difference between the two groups on instrumental dependency is more easy to understand. As noted previously, the *metapelet* in *Kibbutz* S encourages and demands more independence on the part of the children than the *metapelet* of *Kibbutz* M. She urges them to do things by themselves rather than to ask for help. She often refuses help when the children ask for it, telling them instead to try to do whatever they wanted by themselves.

Dependency toward children. There is a general tendency for the amount of affection and dependency expressed toward children to increase with age. That is, the children in the youngest groups showed affection to other children less frequently than the children in the older groups. There were differences between the two *kibbutzim* only in the middle groups. The children of the middle group in *Kibbutz* M showed more affection and dependency toward other children than the similar age children in *Kibbutz* S. It is not clear why this difference should appear in the middle groups and not in the oldest groups. When the amount of affection expressed toward children was compared with the amount of affection expressed toward the *metapelet*, there were no differences found except in the middle group in *Kibbutz* S. Here the children expressed affection toward the *metapelet* significantly more frequently than toward other children. We would expect to find differences in the youngest groups, but this is not the case.

Relationship Between Dependency and Aggression

In terms of overall scores, the children in both *kibbutzim* expressed significantly more dependent than aggressive responses (t = 6.95). There is a significant negative relationship ($r = -.73$) between the frequency scores of aggression and dependency. That is, the children who are more aggressive are less dependent, and the children who are more dependent are less aggressive. Most studies of nursery school children indicate a general low but positive relationship between dependency and aggression. Lois Murphy (3) in a classical study on sympathy and aggression in young children finds a positive relationship between these two types of responses. In a study reported by Sears *et al.* (5) there tends to be a low positive correlation between these two types of responses. These studies were made on older children than those observed here. Perhaps at this age dependency and aggression are alternative responses to frustration and the children who have been successful in using one or the other type of response tend to persist in its use. If we look at each group separately we note that in all groups, except the middle group in *Kibbutz* M, there is a negative correlation between total aggression scores and total dependency scores. The correlations range from $-.09$ to $-.72$. In the middle group in *Kibbutz* M, there is an unexplained positive correlation ($+.82$) between the frequency of occurrence of these two types of responses.

When analysis is limited to the aggressive and dependent responses expressed toward other children, there is no relationship ($r = +.12$) between the occurrence of each of these responses. In the groups taken separately, however, there is, in four out of the six cases, a high positive correlation (.55–.93) between the expression of aggression and dependency toward children. The children show aggression toward each other with significantly greater frequency than they show dependency ($t = 2.00$), in contrast to the findings with respect to total scores. Not unexpectedly, the children thus tend to respond to other children more frequently with aggression and toward adults with dependency. In the groups considered separately, the children in the youngest group in each *kibbutz* are more frequently aggressive toward one another than they are affectionate, whereas in the two older groups the frequency of these two responses does not differ significantly . At the younger ages, the children are probably more affectionate toward adults as the source of their main satisfactions and care, and are aggressive toward other children since other children do not satisfy their needs; if anything, their peers are probably more likely to frustrate and interfere with need satisfaction. The process of learning to be affectionate toward other children is presumably a more complicated process involving the development of identification with one's peers and of capacities and skills of helping one another out.

SUMMARY AND CONCLUSIONS

An exploratory study of the social development of children of toddler age growing up in *kibbutzim* has been described. Since these children live together in groups practically from birth, the setting is propitious for studying the extent and limits of social interaction among very young children. A secondary point of emphasis was the role of the *metapelet* (children's caretaker) and her relationship with the children.

The sample consisted of three groups of children from each of two *kibbutzim*. The children ranged in age from 19 to 38 months at the time the study was begun. Observations were made on the children for a period of six months. Systematic observations were made on the individual children, general observations were made on each group, and discussions were held with the *metaplot* and educators.

The first part of the report deals with a description of the daily routine of the children. Differences between the *kibbutzim* in degree of deliberate educational stimulation and of freedom vs. structure in the setting were seen as having consequences in the relative amount of time the *metaplot* from the two *kibbutzim* spend with the children.

Accompanying differences in the behavior of the children in the two *kibbutzim* were noted mainly in the youngest groups. The children in the youngest group in *Kibbutz* S showed more dependency responses toward adults, whereas the children in *Kibbutz* M tended to suck their thumbs more often and to cry more frequently. The youngest children in *Kibbutz* S seemed to develop at a faster rate in terms of being more aware of their surroundings, more aware of adults, playing more, and speaking earlier. These differences tended to level off in the older groups where no differences of this sort were noted. The children in *Kibbutz* S tended to verbalize group identification and being "big" more than the children in *Kibbutz* M, but no differences were noted in other behavioral indices of these factors except that in *Kibbutz* M the children in the oldest group had higher instrumental dependency scores than the children in *Kibbutz* S.

The role of the *metapelet* is discussed in some detail, in terms of her role as housekeeper, caretaker, and socializer of the children.

The main body of the study is concentrated on the relations among the children. There is a strong group identification, frequent use of the terms "we" and "ours" as against "theirs." Competition exists between groups rather than within groups. Each group defends its members from attack by other groups. There is not a very high degree of social participation among the children, especially in the youngest groups, but the children always play in the presence of the others, and the absence of any child is noted by the other children. The children tend to control each other's behavior by telling each other how to behave. There is a tendency to conformity to group standards. Problems of discipline are negligible. Treatment of sex play and masturbation is very permissive. Toilet training and eating habits seem to be learned with a minimum of conflict.

There are wide individual differences in the frequency of aggressive responses. The frequency of dependent responses is significantly higher than the frequency of aggressive responses, and children who are aggressive tend not to be dependent and vice versa. Children tend to be more aggressive than affectionate toward each other, especially in the youngest groups, and to be as frequently affectionate as aggressive in the older groups. There is a positive relationship between aggression and affection expressed toward other children.

The results of this study indicate, then, that group identification, sharing, and group control of individual behavior can be learned by very young children under conditions of group living. Caution is in order in generalizing from these results in view of the lack of comparative data on similar age children growing up in private families, and of the fact that children in only two *kibbutzim* were studied.

REFERENCES

1. FAIGIN, HELEN. The frequency and types of rules used by preschool children in their social interaction. Unpublished master's thesis, State Univer. of Iowa, 1948.
2. IRVINE, ELIZABETH E. Observations on the aims and methods of child rearing in communal settlements in Israel. *Hum. Relat.*, 1952, **5**, 247–275.
3. MURPHY, LOIS B., *Social behavior and child personality*. New York: Columbia Univer. Press, 1937.
4. PIAGET, J. *The moral judgement of the child*. New York: Harcourt, Brace, 1932.
5. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.
6. SPIRO, M. E. Education in a communal village in Israel. *Amer. J. of Orthopsychiat.*, 1955, **25**, 283–293.
7. SPIRO, M. E. *Kibbutz: Venture in Utopia*, Cambridge, Mass: Harvard Univer. Press, 1956.

Received April 27, 1957.

CRITIQUE AND NOTES

TWO DETERMINANTS OF CONFORMITY: AGE OF INDEPENDENCE TRAINING AND n ACHIEVEMENT[1]

ALLEN M. KREBS
University of Michigan

McClelland (2) has derived a measure of achievement goal learning, n Achievement, and has demonstrated it to be related to conformity. Winterbottom (4) constructed a childhood training schedule to measure the age of self-reliance training; she reported it to be related to a measure similar to McClelland's n Achievement. There has been no attempt made to determine the relationship of Winterbottom's measure to conformity behavior nor the extent to which *both* Winterbottom's scale and McClelland's n Achievement are capable of discriminating between high and low conformers. The formulations of Riesman (3) suggest that both factors in relatively strong operation would render individuals less prone to change their opinions because of external social pressure.

Three hypotheses were derived in an attempt to determine the importance of the two measures:

I. The more intense the orientation toward generalized achievement goals (n Achievement) the more resistant individuals are to opinion change.

II. The earlier the occurrence of independence training during childhood, the more resistant individuals are to opinion change.

III. n Achievement and independence training operate in a synergistic fashion i.e., individuals with low n Achievement and late independence training conform to a greater extent than under any other combination of the two factors.

METHOD

Thirty-six subjects (*S*s), undergraduate males at the University of North Carolina, were employed in this experiment. n Achievement data were collected in the prescribed manner (2); the measure of age of independence training was derived by modifying Winterbottom's questionnaire (originally designed for use with children) to serve as a self report for *S*s of college age.

The following examples of Winterbottom questionnaire items discriminated well between early and late independence training:

——— ——— To show pride in your ability to do things well.

——— ——— To try hard things for yourself without asking for help.

The first space was to be checked if *S* felt that the item was pertinent to his parents' demands; in the second space *S* placed an estimate of the age at which the behavior was demanded of him.

In the situation designed to measure conformity, two *S*s were instructed to remember as many objects as possible of those that were to appear on five successive presentations of a set of two slides. The objects on one slide differed from those on the other—one slide purported to represent objects found inside one town, the other to represent objects found within another town in a context of military intelligence. Since there were a great number of objects on the slides and since the duration of exposure was short, considerable ambiguity existed concerning the precise nature of the objects presented in either. *S*s' scores for conformity were based on the number of objects each *S* was persuaded to indicate had occurred on the "Town A" slide when in fact they had been presented on the "Town B" slide, or vice versa.

Each *S* checked a list containing the names of various objects and indicated in a space next to each object whether it was in Place A or Place B. The *E* then substituted bogus lists for the genuine ones and returned these to the *S*s. Each bogus list purported to represent the responses of the other *S*, and disconfirmed four items of his own list. Conformity scores could range from 0 to 4, depending upon the extent to which *S* accepted the bogus report.

TABLE 1
SUMMARY OF CONFORMITY, n ACHIEVEMENT, AND INDEPENDENCE TRAINING SCORES FOR THE FOUR GROUPS OF SUBJECTS

Group	Conformity		n Achievement		Independence Training	
	M	*SD*	M	*SD*	M	*SD*
Early Independence training High n Achievement	.600	1.075	8.7	2.28	6.92	.169
Early Independence training Low n Achievement	.750	.887	2.2	2.11	6.87	.433
Late Independence training High n Achievement	.875	.991	8.9	4.65	8.62	.679
Late Independence training Low n Achievement	2.40	1.350	3.3	1.90	10.01	2.236

---

[1] This study was conducted in part by the Organization Research Group of the Institute for Research in Social Science, University of North Carolina. The activities of the Organization Research Group are supported jointly by the Ford Foundation and the Office of Naval Research [Nonr-855(04)]. The research is based upon the author's Master's thesis in the Department of Psychology at the University of North Carolina conducted under the direction of Dr. June Chance and with the gratefully appreciated assistance of Drs. deCharms, Thibaut, and Mellinger.

Depending upon whether the *S*s' scores for age of independence training and for n Achievement fell above or below the median values of the two measures, they were assigned to one of four cells: early or late independence training in combination with high or low n Achievement. Presented in Table 1 are summary descriptive statistics of the Winterbottom, n Achievement, and conformity measures. An analysis of variance was employed to determine the significance of differences in conformity among cells.

A *q* test, discussed by Dixon and Massey (1, p. 152, 169), was carried out to determine whether the low n Achievement, late independence training cell indicated greater conformity than the other cells combined proportionately, i.e., weighted one third each.

RESULTS

Results of the analysis of variance of the conformity scores are presented in Table 2. A biserial correlation of .214 was found between measures of n Achievement and independence training. This relationship is not statistically significant ($p > .10$), in contrast to expectations derived from McClelland's formulations.

The significant relationship demonstrated between n Achievement and conformity serves as replication of McClelland's earlier findings. The relationship indicated between independence training and conformity substantiates the second hypothesis.

TABLE 2

ANALYSIS OF VARIANCE OF CONFORMITY SCORES CLASSIFIED BY THE VARIABLES n ACHIEVEMENT AND AGE OF INDEPENDENCE TRAINING

Source	Sum of Squares	*df*	*F*	*p*
Independence training	.9264	1	6.728	<.05
n Achievement	.7013	1	5.093	<.05
Interaction	.4727	1	3.433	N.S.
Error	4.4072	32		

Hypothesis III was supported. The *q* test indicated that the greatest conformity occurred under conditions of low n Achievement, late independence training ($p < .05$). n Achievement and relative age of independence training in concurrent operation seem thus to have a synergistic effect regarding conformity behavior. When an individual falls in the lower half of both measures, he tends to conform to a greater extent than when in the lower half of either measure alone.

REFERENCES

1. DIXON, W. J., & MASSEY, F. J. JR. *Introduction to statistical analysis.* (2nd ed.) New York: McGraw-Hill, 1957.
2. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
3. RIESMAN, D., with GLAZER, N. & DENNEY, R. *The lonely crowd.* New York: Doubleday, 1955.
4. WINTERBOTTOM, M. R. The relation of childhood training in independence to achievement motivation. Univer. of Michigan microfilms, 1953, Publ. No. 5113. (Abstract)

Received February 15, 1957.

THE SCALE VALUES OF STATEMENTS OF ATTITUDE TOWARD THE LAW AFTER A 25-YEAR INTERVAL

JOHN E. MILHOLLAND

University of Michigan

THE VALIDITY of any measurement depends in part on the constancy of the measuring instrument. In the domain of attitudes, measurement of change or of comparative standing of groups, requires that the scale mean the same when used at different times and/or with different groups. No direct test of this requirement is available, of course, but an indirect approach is illustrated by the data reported here.

In May of 1956, 37 students in a graduate class in psychological statistics were assigned an exercise in scaling by the Method of Successive Intervals (1). The attitude statements were taken from the Scale of Attitude Toward the Law (4) constructed by Daniel Katz at Chicago in 1930. As the result of an afterthought, the scale values of the items on the two occasions were compared. The product-moment coefficients of correlation were .98 for Form A of Katz's scale, .99 for Form B. There are 20 statements in each form of the scale.

These high correlations are particularly interesting because the 1956 procedures did not duplicate those of 1930. Katz used the Method of Equal-Appearing Intervals (5), rather than Successive Intervals; and 11 piles, whereas the 1956 scaling was done with 9 piles. The 1956 data were re-analyzed[1] by the Method of Equal-Appearing Intervals, and the correlations between item scale values were again .98 for Form A, .99 for Form B. The correlation between the two solutions for the 1956 data was .99 for each form.

---

[1] Dr. Sam Komorita suggested the re-analysis and carried it out.

Apparently, the 1956 perception of a University of Michigan graduate student as to favorableness of a statement toward law enforcement is not different, in a statistical sense, from that of a 1930 University of Chicago graduate student. Also, the close correspondence between solutions by the Method of Successive Intervals and the Method of Equal-Appearing Intervals makes it seem hardly justifiable to use the more elaborate procedures of the former. These generalizations, however, might not apply to other kinds of attitudes—it may be that the scale of Attitude Toward the Law is peculiarly stable. Hovland and Sherif (3), however, report a correlation of .96 between values of Attitude Toward the Negro (2) items scaled about 20 years apart.

REFERENCES

1. EDWARDS, A. L. Psychological scaling by means of successive intervals. *Psychometric laboratory report.* Univer. of Chicago, 1951, No. 69.
2. HINCKLEY, E. D. Attitude toward the Negro. In L. L. Thurstone (Ed.), *The measurement of social attitudes.* Univer. of Chicago, 1930.
3. HOVLAND, C. I., & SHERIF, MUZAFER. Judgmental phenomena and scales of attitude measurement: item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822–832.
4. KATZ, DANIEL. Attitude toward the law. In L. L. Thurstone (Ed.), *The measurement of social attitudes.* Univer. of Chicago, 1931.
5. THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude.* Univer. of Chicago, 1929.

Received December 17, 1956.

MOTIVATIONAL FACTORS IN THE VIOLATION OF A PROHIBITION

DON L. KIMBRELL AND ROBERT R. BLAKE

University of Texas

CONFORMING to a sign prohibition has been shown to be greater when the sign is "strong" than when it is "weak" in a study by Freed, Chandler, Blake, and Mouton (1). Secondly, more people conform when others conform to a restriction and fewer with knowledge of violations by others (1). Grosser, Polansky, and Lippitt have similar results for violation of a command (2). The power to change the frequency of compliance has, in addition, been shown to be determined partly by "who the others are," with Lefkowitz, Blake, and Mouton (3) investigating this factor in traffic light violations. The present paper extends previous work by relating strength of motivation to act against a prohibition to the social properties of the prohibition situation.

EXPERIMENTAL DESIGN

The experiment was presented to 90 college *S*s as a study of taste preference.

Description of the Test Situation

Two *S*s, an accomplice and an informed *S*, came to the laboratory at the same time. After eating crackers, *S*s were told that there would be one other part of the experiment in which they would participate. Then they were requested to step into the hall to wait until *E* called them. The period was about three minutes. A stimulus sign was placed above the drinking fountain in the hall. The printed sign read, "Do not use this fountain." During the waiting period, the informed *S* either conformed to the sign, or violated it by drinking. For control conditions he was not present.

Induction of Thirst

Three degrees of thirst were induced. Intense thirst was created by feeding six crackers, three treated with Mexican hot sauce and belladonna. Three untreated crackers were used to produce a weak thirst. A puzzle problem was employed for evaluating frequency of drinking when thirst was not manipulated.

RESULTS

Violation with Degrees of Thirst

Inspection of row totals in Table 1 shows that violation increases with the conditions employed to increase thirst. Whereas none of the 20 *S*s under the no thirst condition violated, 16 of 20,

TABLE 1

CRITICAL *S*S' RESPONSES TO THE REACTIONS OF THE INFORMED PARTICIPANT AND TO VARIOUS DEGREES OF INDUCED THIRST

Degree of Thirst	Informed *S*				Total by Thirst	
	Drinks		Does Not Drink			
	Critical *S*		Critical *S*			
	Drinks	Does Not Drink	Drinks	Does Not Drink	Drinks	Does Not Drink
Strong	9	1	7	3	16	4
Weak	5	5	0	10	5	15
Non induced	0	10	0	10	0	20
Total by informed *S*	14	16	7	23		

or 80 per cent, violated under the strong thirst condition. The weak thirst condition, with 5 violations among 20 *S*s, or 25 per cent, elicited a larger number of violations than the no thirst condition. The χ^2 of 25.65 for these differences (2 *df*) is significant far beyond the .01 level. Yates' correction for continuity was applied to χ^2 values.

Violation with Social Properties of Background

Column totals demonstrate that more violations occur when the *S* sees another person violating than when he sees another person conforming to the prohibition. Fourteen of 30 responses were violations when the confederate violated, whereas only seven of 30 were violations when the confederate conformed. The χ^2 for this difference is 3.62 and, with 1 *df*, approaches significance at the .05 level. Computed for the two thirst conditions, those under which violations would be expected to occur, the χ^2 is 4.80 and, with 1 *df*, approaches significance at the .01 level. The χ^2 of 7.20 (1 *df*) for the weak thirst condition, the condition in which *S* is neither intensely thirsty nor not thirsty, is significant beyond the .01 level. Under this condition, the action of the confederate has maximum influence on the *S*s' behavior.

Violation as a Joint Function of Social Properties and Degree of Thirst

The χ^2 of 1.63 for the interaction between social background and strength motivation is not significant, indicating that there is no complex interrelationship between the two variables. The higher frequency of violations associated with increased thirst is independent of higher frequency associated with the behavior of the other person in the situation. Both factors, therefore, are important determinants of readiness to act against a prohibition.

SUMMARY

Ninety students who served as *S*s understood that they were participating in an experiment in the field of taste, when the factor under investigation was behavior relative to a prohibition. The drive varied was thirst, the prohibited action was drinking water from a fountain, and the social background was perceived conformity or violation of the prohibition by another person. Behavior produced by a physiological drive of given strength is shown to be determined partly by the properties of the drive and partly by other properties of the situation which are social in character and which are not intrinsically related to the drive.

REFERENCES

1. FREED, A. M., CHANDLER, P. J., MOUTON, JANE S., & BLAKE, R. R. Stimulus and background factors in sign violation. *J. Pers.*, 1955, **23**, 499.
2. GROSSER, D., POLANSKY, N., & LIPPITT, R. A laboratory study of behavioral contagion. *Hum. Relat.*, 1951, **4**, 115.
3. LEFKOWITZ, M., Blake, R. R., & MOUTON, JANE S. Status factors in pedestrian violation of traffic signals. *J. abnorm. soc. Psychol.*, 1955, **51**, 704–705.

Received March 7, 1957.

THE ACHIEVEMENT MOTIVE AND TASK PERFORMANCE: A REPLICATION

ROBERT C. BIRNEY

Amherst College

THE findings reported here grew out of an effort to replicate previous research on the achievement motive (n Ach), in which a positive relationship had been reported between n Ach and simple speed tasks. Lowell (5) demonstrated a positive relationship between n Ach scores and a speed test of simple addition, while French (3) showed the same relationship using a digit-substitution task with a population of officer candidates. The intent of the study was to repeat as precisely as possible the original conditions of Lowell. The difficulties encountered in effecting this replication seem worthy of report.

METHOD

Four groups were used. Group 1 consisted of 24 male *S*s who volunteered from the introductory class in psychology in which the investigator was an instructor. The investigator served as *E*. Group 2 consisted of 21 male *S*s from the investigator's class, with another faculty member well known to the students as *E*.[1] For neither of these groups were Lowell's results confirmed. A re-examination of his original work brought out the fact that Lowell at the time of his study had been a graduate student experimenting with *S*s from a sister institution. Similarly, French was apparently unknown to her *S*s before the experiment. Accordingly, it seemed that a closer replication of Lowell's conditions demanded a student *E*. Group 3 consisted of 21 male *S*s from one of the writer's classes, but the *E* was an honor student[2] in psychology. Group 4 consisted of 22 male *S*s from the introductory course, again with the

---

[1] The writer is indebted to Professor Robert Grose for his assistance as *E* for this group.

[2] The writer is indebted to Gary Kirshbaum for his assistance as *E* for Groups 3 and 4.

TABLE 1

EFFECT OF EXPERIMENTER, INSTRUCTIONS, AND ACHIEVEMENT MOTIVE ON ADDITIONS AND DIGIT-SUBSTITUTION TASKS (Lowell's original findings included for comparison)

Group	Faculty *E*		Student *E*		Lowell (5) (*N* = 40)
	Group 1 (*N* = 24)	Group 2 (*N* = 21)	Group 3 (*N* = 21)	Group 4 (*N* = 22)	
Mean Additions					
High n Ach	146.5	133.3	157.8[a]	149.1	157.4[a]
N	14	11	11	9	19
Low n Ach	149.0	136.1	130.7	127.8	130.6
N	10	10	10	12	21
Mean Digit-Substitution					
High n Ach	260.6	218.9	281.0[a]	255.2[a]	
N	14	11	11	9	
Low n Ach	266.0	218.0	234.1	223.7	
N	10	10	10	12	
Total Group Means					
n Ach	5.54[b]	4.19[c]	+.57	−.27	
Additions	147.5	134.5	145.0	137.0	143.3
Digit-Substitution	262.7	218.5[d]	258.6	236.6	

[a] High n Ach performance reliably greater than Low's at .05 level.

[b] Reliably greater than Groups 3 and 4 at the .01 level.

[c] Reliably greater than Group 3 at the .05 level, Group 4 at the .01 level.

[d] Reliably less than Groups 1 and 3 at the .05 level.

same student acting as *E* but with a change in performance task instructions.

All groups were given the n Ach TAT in booklet form with standard instructions (6, p. 98).[3] After the TAT, the Additions task was administered followed by the Digit-Substitution task. The Additions task (summing three two-digit numbers) ran for ten minutes, one minute per page of twenty-five problems. The Digit-Substitution task ran for eight minutes, one minute per page of seventy letter-number combinations.

Groups 1, 2, and 3 received the following achievement orientation instructions after writing the stories and before attempting the tasks:

> The second portion of this session is devoted to performance on two tasks which are enlarged versions of two typical intelligence test items. Both are speed tests, and both represent abilities which have been consistently tested in IQ tests. My purpose is to compare your performance on them, and to take the first steps in research to see whether or not longer samples of these behaviors might be more diagnostic than the shorter versions now in use. Put your name on the booklets when you receive them. . . . Before we pass the first booklet out, however, let me say that unless you do your very best on these tasks I cannot make good use of the results, so please give it all you have.

Group 4 was told to read the printed instructions, which described the task and its timing, and then they were set to work ("neutral" instructions). "Neutral" instructions were used with Group 4 to explore the nature of the n Ach-performance task relationships under minimal cue conditions.

RESULTS

The results for all groups appear in Table 1. A direct comparison with Lowell's results is given. Such a comparison is impossible for French's data since her results are reported as mean improvement scores between control and experimental sessions. Two main findings appear. First, the mean n Ach score is significantly greater with the faculty *E*s than with the student *E*. Second, Group 3, which is the group closest to Lowell's original conditions, provides a precise replication of his results. In Group 4 high and low n Ach differ reliably on the Digit-Substitution task but not on the Additions. A comparison of total group means for the tasks show no group differences on the Additions task, but Group 2 drops reliably below the other groups on the Digit-Substitution task.

DISCUSSION

In regard to the effects of the *E*s upon the n Ach score, the faculty persons with whom the *S*s had associated in classroom (achievement) settings appear to have had the effect of arousing imagery in *S*s who were not motivated to perform well on the tasks. This interpretation is consistent with McClelland's theory of imagery determination, that the presence of achievement imagery in a story is due in part to the strength of the achievement cues in the story writing situation (6, p. 194). Had the higher n Ach scores for Groups 1 and 2 reflected greater motivation we would expect these two groups to outperform Groups 3 and 4 on the tasks, but they did not do so. In fact, the over-all level of performance on the tasks is comparable across the four groups, and data for the Additions task are consistent with the findings of Lowell.

The results for Group 3 constitutes the closest *E* replication of Lowell's findings. Here a student *E* arouses far less achievement imagery, but the occurrence of such imagery predicts performance on the tasks successfully. Previous research relating n Ach to performance tasks of the type used here has not involved the use of a faculty *E* having close association with the *S*s. The possibility is thus suggested that an *E* effect has been present in many previous studies without sufficient appreciation of its role.

Two further aspects of the data warrant attention. First, Group 4 shows no drop in mean task performance despite "task" as opposed to "achievement arousal" instructions before the tasks. Similar findings have been reported by French (3) and Johnston (4). Apparently, for some populations the difference between aroused and task instruction conditions has little effect on performance. Second, Group 2 shows a considerably lower mean performance on the Digit-Substitution task despite a substantial level of achievement imagery. This group was run at 11:00 A.M. on a sunny Saturday morning in May. Analysis of their Digit-Substitution task performance, which came toward the end of the hour-long test session, shows a mean performance of 8.6 letters below Group 1 on the first trial, and 24.4 letters below Group 1 on the last four trials, the sharp drop toward the end of the hour accounting for the lower over-all mean.

---

[3] The pictures used were: TAT 7BM, TAT 8 BM, the inventors (see p 100, 6), the printers, (used by Lowell), and the college class, #21, and office, #17, from Birney (1).

Two alternative explanations for the general findings were explored. The difference in n Ach scores from the faculty *E*s to the student *E* might be due to shift in scorer set, since these data were being processed as they were obtained. This possibility was checked by selecting seven *S*s from each group, matched within one point on total n Ach score, scrambling the stories, and scoring them without knowledge of their source. Scorer reliability for the scores of the 28 *S*s is +.86 (rho), indicating no large or systematic shifts in scorer set.

A second possibility is that the achievement imagery obtained by the faculty *E*s contains indices of anxiety or fear of failure which in turn might be related to the performance measures. Following Clark (2), an analysis was made of the affective signs (positive or negative) of the various categories of the n Ach scoring system. Clark's work suggested that persons characterized by fear of failure and achievement anxiety might tend to write stories dominated by negative categories. A careful analysis of the stories of Groups 1 and 2 show no category distributions that might imply greater anxiety.

The present results thus imply, first, that subsequent research in this area should give careful attention to the experimenter-subject relationship, and second, that this relationship may profitably be taken as an independent variable.

SUMMARY

Efforts to repeat the work of Lowell and French showing a positive relationship between achievement motivation and striving behavior are described. Two groups in which the *E* was a faculty member known to the students failed to reproduce the earlier results. Substitution of a student *E* resulted in good replication of the findings of Lowell and of French.

REFERENCES

1. BIRNEY, R. C. Studies on the role of picture cues in projective measurement of achievement motivation. Unpublished doctoral dissertation, Univer. of Michigan, 1955.
2. CLARK, R. A. A factor analytic integration of projective and clinical measures of need achievement. Unpublished master's thesis, Wesleyan Univer., 1948.
3. FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.
4. JOHNSTON, R. A. The effects of achievement imagery on maze-learning performance. *J. Pers.*, 1955, **24**, 145–152.
5. LOWELL, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1951, **33**, 31–40.
6. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

Received March 14, 1957.

TWO COMPONENTS OF ASSUMED SIMILARITY[1]

ELIZABETH ALFERT

University of California

A MEASURE in social perception developed recently by Cronbach *et al.* (2) is the one of Assumed Similarity (AS) which is derived by comparing a person's self description with his prediction of the self description of others. Fiedler (3) has regarded AS as a measure of psychological distance, and has used it in connection with sociometric choices to investigate the structure and dynamics of social groups (e.g., 4). Rudin *et al.* (9) found AS scores to be consistent from one series of items to another and considered Assumed Similarity a measurable personality trait in which consistent individual differences exist. Cronbach (1) has called Assumed Similarity a general mental set almost independent of the psychological content of the items used, for which a global index would provide a satisfactory score.

This investigation is concerned with the processes involved in interpersonal perception. The possibility has been considered that the attribution of one's traits to others may not be a homogeneous mental set, but may involve two different predispositions: a predisposition to attribute traits in which ideal self and self coincide (ideal-congruent traits), and one to attribute traits in which ideal and self differ (ideal-discrepant traits).

---

[1] The author is greatly indebted to Dr. Egerton L. Ballachey who made this study possible by permitting the collection of data in his classes, and who has contributed much valuable advice and criticism.

If this could be established, Assumed Similarity would have to be considered a complex mental set, and an individual's global Assumed Similarity score would be a function of the relative number of ideal-congruent and ideal-discrepant items in the sample of items used.

In order to test whether Assumed Similarity scores are dependent on item content, the hypothesis is examined that *ideal-congruent traits and ideal-discrepant traits are attributed to others to a different degree*, by comparing individuals' mean Assumed Similarity scores for ideal-congruent traits (AS$_c$) with their mean Assumed Similarity scores for ideal-discrepant traits (AS$_d$). It may also be of interest to see whether individuals' AS$_c$ scores can be predicted from their AS$_d$ scores, and the relationship between AS$_c$ and AS$_d$ scores will, therefore, be investigated.

A person who attributes his ideal-congruent and ideal-discrepant traits to others in equal degree would have an AS$_d$/AS$_c$ ratio of 1.00. A ratio smaller than 1.00 indicates that ideal-congruent traits are attributed to others to a greater extent than ideal-discrepant traits, while a ratio of more than 1.00 means that ideal-discrepant traits are attributed more than ideal-congruent traits. The present study also seeks to throw light on whether an individual's AS$_d$/AS$_c$ ratio is indicative of a fairly consistent attitude and whether it is determined to any extent by his self satisfaction (the number of ideal-congruent traits). A further hypothesis, thus, holds that *there is a significant correlation between the amount of ideal-self congruency and the ratio of* AS$_d$/AS$_c$.

METHOD

Since this study is concerned with the influence of processes in the perceiver rather than with the accuracy of the perception of the behavior of the person perceived, unstructured stimuli, namely strangers were used. Gage (5) has shown that predicting the responses of strangers turned out to be less accurate than predicting the average. Judging strangers may be a more life-like situation and may thus reveal more of the motivational factors of the perceiver.

Two experimental groups of mixed sex were used: Group A, an adult extension class group of 27 people, and Group B, an undergraduate class group of 13 students. The students were unacquainted with each other at the time of the study. The stimulus persons to be described were members of the group, homogeneous as to race and social class, and they alternated in sex. Six stimulus persons were used for Group A, twelve persons for Group B.

The trait list used for Group B was constructed from Murray's set of personality variables (7), the 20 needs and five of the general attributes, choosing a trait name or attitude pertaining to a certain need and pairing it with an adjective of opposite meaning. A similar but less systematic list of 20 opposite traits had been used for Group A. Item pairs such as forceful—gentle, light-hearted—serious, spontaneous—self-controlled are examples of pairs used.

Each student in Group A and Group B was asked: (*a*) to check the trait list of opposite pairs, checking in each trait pair the one which would describe you better; (*b*) to check the list, according to your guess, as several specific individuals would check it for themselves; (*c*) to check the list descriptive of your self ideal, how you would ideally like to be.

In Group B, an additional task was added between steps 1 and 2. Each student was asked to check the list so as to describe, according to his guess, other specific individuals in the class.

Procedure for Scoring

AS has usually been scored by finding the difference between *S*'s self-description and *S*'s prediction of *O*'s self-description, but since a dichotomy is involved here, AS is expressed in terms of correspondences between *S*'s self-description and his prediction of *O*'s self-description.

1. AS$_g$, the global Assumed Similarity score was computed for each individual by summating correspondences between description of self and the predicted trait of the stimulus person for all items and for all persons, dividing by the number of stimulus persons and by the number of items.

2. An AS$_c$ score for each individual was calculated by summating correspondences between self and the predicted trait of the stimulus person for all persons and for all ideal-congruent items, dividing by the number of stimulus persons and by the number of ideal-congruent items.

3. An AS$_d$ score for each individual was computed by summating correspondences between self and the predicted trait of the stimulus person for all persons and for all ideal-discrepant traits, dividing by the number of stimulus persons and by the number of ideal-discrepant items. AS$_g$, AS$_c$ and AS$_d$ based on the description of *O* rather than on the prediction of *O*'s responses are similarly calculated.

AS$_c$ can be considered as the AS score for the average ideal-congruent trait or the percentage of *O*s to whom the average ideal-congruent trait is attributed, while AS$_d$ is the AS score for the average ideal-discrepant trait.

RESULTS

Ideal-self congruency in Group A ranged from 40 to 85 per cent, with a mean of 62 per cent and in Group B from 36 to 84 per cent, with a mean of 70 per cent. Agreement as to which member of the trait pair was the ideal varied from unanimous acceptance of one to about equally divided opinions, depending on the item pair.

Hypothesis 1. Ideal-congruent traits are attributed to others to a different extent than ideal-discrepant traits. Mean scores for AS$_c$ and AS$_d$ are given in Table 1. The difference between individuals' AS$_c$ and AS$_d$ was calculated and the significance of the mean difference tested. The difference between AS$_c$ and AS$_d$ is significant beyond the .001 level of confidence, in the prediction of O's self-description as well as in the description of O. (No significant difference at the .05 level was found for Group B between AS scores based on prediction and AS scores based on description.) An individual's global AS falls somewhere between his AS_c and AS_d score, approaching AS_c with greater ideal-self congruency. It thus seems evident that the psychological content of the item influences the extent to which similarity is assumed.

The possible influence of the general desirability of a trait (as determined by the opinion of the group) on the attribution of one's traits to others was also checked. Item pairs in Group A and B have been divided into two sets, 10 "extreme" item pairs for which the group was unanimous or almost unanimous as to which of the pair was the desirable trait, and 10 "neutral" item pairs for which less than two thirds of the group agreed as to the desirable trait. It was found that the average number of ideal-congruent items out of the ten was six in Group A for the extreme item pairs as well as for the neutral ones, and seven in Group B in extreme as well as in neutral items. Table 2 shows that AS rises with the general desirability of the trait. To some extent, though, the traits considered most unanimously as desirable may also be the ones among the ideal-congruent traits which are subjectively considered the most desirable ones. When the significance of the mean difference between AS_c and AS_d was tested in the neutral trait pairs, it was seen that the difference between the attribution of ideal-congruent and ideal-discrepant traits was still significant, $t = 3.15$ ($p < .01$) for Group A, and $t = 2.30$ ($p < .05$) for Group B.

A rank order coefficient of .08 between AS_c and AS_d was found for Group A and a rank order coefficient of .16 for Group B. The correlations are not significant at the .05 level. Though this finding does not prove that the variables are independent, it suggests that whatever association exists between them is weak.

Hypothesis 2. The AS_d/AS_c ratio was not significantly correlated with the amount of ideal-self congruency. A rank order coefficient of .10 between AS_d/AS_c and ideal-self congruency was found for Group A and a rank order coefficient of .18 for Group B. Plotted data show that people with high, medium, or low ideal-self congruency were equally likely to have a certain AS_d/AS_c ratio.

TABLE 1

MEAN ASSUMED SIMILARITY SCORES FOR IDEAL-CONGRUENT AND IDEAL-DISCREPANT TRAITS

Group and Task	N	AS_c	AS_d	t	Level of Significance
Group A prediction	27	.64	.46	4.8	$p < .001$
Group B prediction	13	.74	.47	4.9	$p < .001$
Group B description	13	.71	.46	5.2	$p < .001$

TABLE 2

MEAN ASSUMED SIMILARITY SCORES FOR 10 EXTREME AND 10 NEUTRAL TRAIT PAIRS

Category of Trait	AS (Group A)	AS (Group B)
Ideal-congruent, unanimously considered desirable traits	.65	.82
Ideal-congruent traits of neutral items	.60	.65
Ideal-discrepant traits of neutral items	.49	.53
Ideal-discrepant, unanimously considered undesirable traits	.40	.33

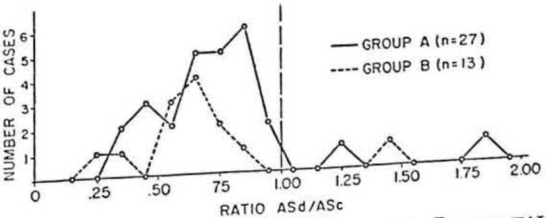


FIG. 1. FREQUENCY DISTRIBUTION OF INDIVIDUAL RATIOS OF AS_d/AS_c

In order to examine whether the AS_d/AS_c ratio as well as AS_c and AS_d are indicative of fairly consistent attitudes within the individual, some of the reliabilities were checked. In Group A, split-half reliability was .64 for ideal-congruent items, .63 for ideal-discrepant items, across-person reliability was .57 for AS_c, .54 for AS_d and .58 for AS_d/AS_c. In Group B, slightly different item pairs were used, the number of item pairs increased from 20 to 25, and the number of stimulus persons increased from six to twelve. Across-item reliability was .86 for ideal-congruent items and .45 for ideal-discrepant items. Among the 25 items, the mean number of ideal-discrepant items was seven so that split-half reliability in the latter case could not have been expected to be very high. Across-person reliability for Group B was .97 for AS_c, .86 for AS_d and .93 for the AS_d/AS_c ratio. It seems that the AS_d/AS_c ratio is indicative of fairly consistent attitudes. Fig. 1 shows that the AS_d/AS_c distributions in groups A and B are of very similar appearance, with a mean of .72 for Group A, and .64 for Group B. Only three individuals out of 40 had ratios greater than 1.

DISCUSSION

There appears to be a general tendency for persons to perceive their ideal-congruent traits in others to a greater extent than their ideal-discrepant traits. When individuals face unknown members of their group, they have a definite opinion as to how they compare to them and in general they feel themselves more alone in traits they consider as shortcomings.

Results based on the prediction of *O*'s responses alone might have been interpreted as *S*'s belief that *O* would be unlikely to attribute undesirable traits to himself, but since very similar results are derived regardless of whether AS is based on the description of *O* or on the prediction of *O*'s responses, it seems that *S* does not make much distinction between predicting *O*'s responses and perceiving *O*. Both probably are a function of perceiving *O* in some relation to the self image. The way others are perceived depends to a large extent on (*a*) the self image, the amount of ideal-self congruency or self-esteem, (*b*) the assumption that others are similar to oneself, (*c*) the linear relationship between desirability of the trait and the amount of assumption of similarity, and (*d*) the AS_d/AS_c ratio, the person's particular disposition to use different standards for attributing his ideal-congruent and his ideal-discrepant traits.

Since the stimulus is ambiguous, judgment of strangers could be considered as a projective method, and extreme AS_d/AS_c scores at either end might be taken as signs of maladjustment. People who are unable to see others as sharing their short-comings may suffer from feelings of inferiority. On the other hand, the rare cases who are extreme in predominantly seeing their shortcomings in strangers rather than their good qualities, probably use this as ego defense or compensation mechanism. It seems likely that they felt the discrepancies between self and ideal so intensely that they had to compensate for them in order to maintain equilibrium. Klein and Schlesinger (6) have discussed the role of perception for maintaining equilibrium.

Zimmerman (10) showed that discrepancy between self and ideal may not indicate conflict. The present results for Hypothesis 2 indicated that ideal-self congruency does not influence the AS_d/AS_c ratio; it seems that traits discrepant from the ideal can be accepted by the ego, and only if such discrepancies become unacceptable to the ego do people attribute ideal-congruent traits to a lesser extent than ideal-discrepant traits in order to re-establish equilibrium. Rogers (8) sees adjustment related to the correspondence between self and ideal. Considering AS_d/AS_c scores in addition to ideal-self congruency might provide a better method for judging adjustment, since adding this measure would not only indicate a person's ideal-self discrepancies but also show how he reacts to them. It would be interesting to see whether extreme AS_d/AS_c scores are associated with certain maladjustments and whether therapy would revert such ratios towards the central tendency. Maladjustment has often been considered a disturbance in interpersonal relationship; how a person perceives others and how he perceives himself within the framework of his group might be a way of measuring it.

SUMMARY

1. Ideal-congruent traits are attributed to others to a greater extent than ideal-discrepant traits.

2. Assuming similarity for ideal-congruent traits and assuming similarity for ideal-discrepant traits are two separate attitudes and determine the global AS score.

3. Ideal-self congruency does not determine the relationship between attributing ideal-congruent and ideal-discrepant traits to others.

4. Central tendency and clinical implications of extreme cases are discussed.

REFERENCES

1. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assuming similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
2. CRONBACH, L. J., HARTMANN, W., & EHART, M. E. Investigation of the character and properties of Assumed Similarity measures. *Univer. of Illinois Bur. of Res. and Serv.*, 1953, Tech. Rep. No. 7, Contract 6 ori-07135.
3. FIEDLER, F. E. The psychological distance dimension in interpersonal relations. *J. Pers.*, 1953, **22**, 142–150.
4. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381–388.
5. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
6. KLEIN, G. S., & SCHLESINGER, H. J. Where is the perceiver in perceptual theory? *J. Pers.*, 1949, **18**, 32–47.
7. MURRAY, H. A. *Explorations in personality.* New York: Oxford Press, 1938.
8. ROGERS, C. R. The significance of the selfregarding attitudes and perceptions. In M. S. Reymert (Ed.), *Feelings and emotions.* New York. McGraw Hill, 1950.
9. RUDIN, S. A., LAZAR, I., EHART, M. E., & CRONBACH, L. J. Some empirical studies of the reliability of interpersonal perception scores. Tech. Rep. No. 4, Contract N6 ori-07135, Urbana, Ill.: Bur. Res. Serv., Univer. of Illinois, 1952.
10. ZIMMERMAN, H. Self acceptance and its relation to conflict. *J. consult. Psychol.*, 1954, **18**, 447–449.

Received March 21, 1957.

NEUROTICISM, EXTRAVERSION, AND THE PERCEPTION OF THE VERTICAL[1]

RONALD TAFT
University of Western Australia

AND JAMES COVENTRY
New South Wales University of Technology

IN THEIR well-known studies of the relationship between personality factors and various tasks involving the perception of the upright (5), Witkin and his associates conclude that the relevant personality characteristics can "be reviewed under three headings: the nature of the individual's relation to his environment (including other people), the way in which he manages his impulses and strivings, and the kind of conception he has of himself" (p. 467). This present study deals with the relationship between (*a*) a simple test of ability to perceive the upright when, in the absence of a visual frame of reference, the angle of the body itself must be used as a reference, and (*b*) the personality traits Neuroticism and Extraversion-Introversion. These traits were selected by Eysenck (2) as representing the two main factors constituting the personality domain. Extraversion-Introversion appears to be similar to Witkin's first heading, and Neuroticism to his second, and perhaps his third. Let us now look at these in more detail.

Neuroticism. This trait, perhaps better referred to as emotional instability, implies inability to accept or to control one's impulses and strivings and lack of insight leading to anxiety, tension and self-deprecation. These characteristics suggest poor integration of the self, or "weak ego." Witkin's results indicate that persons who are high on these neurotic tendencies tend to be inaccurate in their perception of the upright. Witkin's findings are consistent with a common sense expectation that a neurotic person, deprived of the usual environmental cues upon which everyone depends in his normal life, would be less able to adapt himself to the use of somatic cues only, under the unusual circumstances of these experiments. Therefore, Hypothesis 1 is that *persons scoring high on a test of neuroticism perceive the upright less accurately than persons scoring low in neuroticism.*

Extraversion-Introversion. By extraversion we mean a personality orientation towards one's environment as opposed to social reserve, introspectiveness and self-sufficiency. This dimension is not quite the same as Witkin's "passivity-activity" relation to the environment, but we should expect the extravert to be more dependent on his environment, and therefore more "passive" in Witkin's sense than the introvert. Consequently, extraverts would be relatively unprepared to handle the environment by means of cues emanating from themselves rather than from the environment. Thus, Hypothesis 2 is that *persons scoring high on extraversion perceive the upright less accurately than those scoring high on introversion.*

METHOD

Personality Test

For convenience, a group-administered test, the Cattell 16 PF, was used to test neuroticism and extraversion-introversion: Factor C, "Emotional Stability or Ego-Strength versus General Neuroticism," was used to measure neuroticism and Factor H, "Adventurous Cyclothymic versus Inherent Withdrawn Schizothymic," to measure extraversion-introversion.[2] The test was scored in accordance with the norms published in the manual (1).

Spatial Orientation Apparatus

The equipment, consisting of a tilting chair and a luminous rod, was adapted from that described in (4). The chair was mounted on a length of tubing so that it could be turned in a vertical plane to the horizontal on either side, and the angle of tilt was measured by a protractor at the rear. The subject's position in the chair was kept constant by the use of shoulder pads, head rest, foot slots and a handle in the middle to keep the hands in a fixed position.

The rod was 39 inches long and situated five feet in front of the subject (*S*). From a handle operated behind the chair the rod could be turned in the vertical plane through any angle. It was illuminated by a series of small globes shining through a narrow slit along the center of the rod.

The testing was carried out in a dark room, the only source of light being the illumination from the rod.

Subjects and Experimental Set

The experimenter (*E*) administered the 16 PF to all Psychology I students in one of their regular laboratory

---

[1] This report is based on work done at the University of Western Australia. The authors thank Dr. A. J. Marshall for his assistance in the planning and administration of the perceptual testing.

[2] The traits included in the neurotic pole of Factor C are "lacking in frustration tolerance, changeable, showing general emotionality, evasive, neurotically fatigued, and worrying." On the extraverted pole of Factor H the traits are "gregarious sociability, adventurous, bold, having marked interest in the opposite sex, frivolous, strong artistic or sentimental interests, abundant emotional response."

TABLE 1

MEAN ERROR IN PERCEIVING THE VERTICAL WITH BODY TILTED AT VARIOUS ANGLES

Group	*N*	Tilted Right				Tilted Left			
		28°	42°	90°	Average	28°	42°	90°	Average
High C factor	26	3.7	6.0	10.0	6.6	3.6	6.8	11.0	7.0
Low C factor	21	3.5	6.0	10.2	6.7	2.8	5.3	9.4	5.8
t		—	—	—	—	1.01	1.15	1.01	1.30
High H factor	21	4.3	8.1	11.9	8.2	3.9	6.8	12.2	7.5
Low H factor	26	3.1	4.0	8.3	5.1	2.4	5.3	8.2	5.3
t		1.53	3.15**	2.14*	3.25**	2.30*	1.15	2.89**	2.52*

* Significant at 0.05 level.

** Significant at 0.01 level.

periods. They understood that they were taking this test both as part of their orientation to psychology and as part of their course obligation to act as *S*s for a number of student and staff research projects. No mention was made at this stage of a second part of the research. Unbeknown to the *S*s, only males were used and the tests were scored only on Factors C and H. The "middle" scorers (approximately the middle 50 per cent) on either factor were then discarded. The *E* then approached the remaining 47 men and sought an appointment for their participation in another experiment. There were no refusals. The *S*s were distributed as follows: high emotional stability (score 8–10 on C) 26, low emotional stability (1–5) 21; extraversion (7–10 on H) 21, introversion (1–3) 26. They were distributed in the four quadrants as follows:

High Stable—extravert	12
Low Stable—extravert	12
High Stable—introvert	14
Low Stable—introvert	9

Before taking *S* into the dark room, *E* explained that some work was being done on the perception of the upright, and that he was interested in how well people could judge the upright while they were in different body positions. All *S*s were told that nothing unexpected would happen to them, and were made to feel at ease before entering the dark room.

Procedure

In order to allay anxiety, *S*s were seated and "clamped" in the chair under full lighting conditions. The lights were then turned off until the completion of the testing. The *S*s were instructed as follows:

> I want you now to sit in this chair. You will be tilted at various angles in the vertical plane only. Place both hands on this handle and keep hold of it all the time. You need have no fear of falling as the chair has been specially weighted to counterbalance your body weight. See this rod in front of you? The rod lights up. When I switch it on I want you to tell me whether it is upright or not. By upright, I mean upright from the floor, not in relation to you. If it is upright say "It is upright." If it is not upright say "It is not upright."

Once these instructions were given, *E* seated *S* in the chair and adjusted the head rest and shoulder pads, making sure that *S* was as comfortable as he could be and that he had both hands placed firmly on the handle.

Judgments were then made by the *S* while tilted at the following angles:

(*a*) 0° Upright
(*b*) 28° Right
(*c*) 42° Right
(*d*) 90° Right
(*e*) 28° Left
(*f*) 42° Left
(*g*) 90° Left
(*h*) 0° Upright

In each of these body positions the rod was placed in an initial position 40° to the left of the upright and then adjusted by *E* until *S* was satisfied that it was vertical. Before *S* was tilted to a new angle these judgments were repeated three more times alternating the starting positions between 40° right and 40° left. The rod was moved from the starting positions in small increments, the light on the rod being switched off between each move.

On the average, the total time for each *S* was 30 minutes, with a maximum of 40 minutes.

Scoring

In each case the deviations from the vertical position were read to the nearest degree, disregarding sign. The deviations for the four judgments by each *S* were averaged in order to give a mean error score for every *S* at each of the eight positions.

RESULTS

In no case did the mean error score at 0° reach 1, so 0° is taken as the base for each *S*. The results for the other positions are summarized in Table 1. These results indicate no difference between the emotionally stable and unstable groups, and thus Hypothesis 1 is not confirmed. If anything, the unstable group are superior at the task than the stable. Hypothesis 2 is confirmed; the extraverts are consistently less accurate than the introverts.

When the *S*s falling into the four personality quadrants are compared, the nonneurotic introverts are the most accurate and the nonneurotic extraverts the least. The difference between these two extreme groups is significant at the 0.01 level for both the left and right tilts. The nonneurotic introverts are also significantly more accurate than the neurotic extraverts (on the right tilt only).

DISCUSSION

The results for emotional stability appear to contradict those of Witkin and associates (5) and the common sense expectation that a person who is uncertain and in a state of conflict about himself in relation to his environment will be handicapped in using his own body as a framework for judging the environment. Why, then, did the less stable *S*s perform as well as the more stable? It is impossible to answer this from the present data, but the discrepancy between our expectation and the findings could possibly be due to a mistaken opinion about the meaning of Factor C on the 16 PF test. In the manual for the test (1) the authors identify this factor with Eysenck's General Neuroticism, but it may not correspond with the characteristics described by Witkin and associates as "lack of awareness of inner life, fear of aggressive and sexual impulses, and poor control over these impulses . . . accompanied by . . . considerable anxiety, combined with difficulty in regulating it" (5, p. 468). It is possible that some persons "lacking awareness of their inner life" may appear more emotionally stable on a personal inventory such as the 16 PF than they really are.

Failure to obtain the expected results on neuroticism is unlikely to be explained in terms purely of the type of test used to perceive the upright. The task in our test appears to be psychologically equivalent to Witkin's tilting-room-tilting-chair Series 1-a and 1-b (5); i.e., it is a "field-as-a-whole task" in which the object of perception (the rod) cannot be directly related to the body. In these studies, poor performance on this type of test correlated significantly with a number of measures indicating neuroticism, including interview ratings of the *S*'s "anxiety versus self-assurance."

Apart from the possible misinterpretation of Factor C, we are at a loss to explain our results on neuroticism.

The confirmation of Hypothesis 2 requires little comment. Evidently, Factor H which we identified as extraversion is relevant to the ability to perceive the upright. Following Linton's analysis of the tilting-room-tilting-chair tests (3), we make the suggestion that Factor H is a measure of the "generalized tendency to accept or reject external influence."

SUMMARY

Forty-seven male *S*s were tested on their ability to judge when a luminous rod was vertical, while they were tilted sideways in a chair within a dark room. It was hypothesized that persons scoring higher on tests of neuroticism and extraversion would be less accurate in perceiving the vertical than those scoring low on these traits. The traits were measured by the C and H scales respectively from the Cattell 16 PF questionnaire.

No difference was found between those scoring high or low on neuroticism, but those scoring high on extraversion were significantly less accurate in perceiving the vertical than were those scoring low.

REFERENCES

1. CATTELL, R. B., SAUNDERS, D. R., & STICE, G. F. *The sixteen personality factor questionnaire.* Champaign, Ill. Institute for Personality and Ability Testing, 1950.
2. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
3. LINTON, HARRIET B. Dependence on external influence: correlates in perception, attitudes, and judgement. *J. abnorm. soc. Psychol.*, 1955, **5**, 502–507.
4. WITKIN, H. A., & ASCH, S. E. Studies in space orientation III. Perception of the upright in the absence of a visual field. *J. exp. Psychol.*, 1948, **38**, 603–614.
5. WITKIN, H. A. *et al. Personality through perception.* New York: Harpers, 1954.

Received August 9, 1956.

ON AUTHORITARIANISM AND ACQUIESCENCE: AN ADDED NOTE TO BASS AND MESSICK AND JACKSON

FRED N. KERLINGER
New York University

IN the May 1957 issue of this Journal, Messick and Jackson (2) criticized Bass's treatment and interpretation of the data of his study of authoritarianism and acquiescence (1). In their criticism Messick and Jackson stated that Bass's factor loadings have an incorrect sign. They also change the labels of the two factors Bass "extracted." This of course changes the interpretation of the table radically. I would like to add a bit more discord to this composition by transposing one of its themes to another key.

Neither Bass nor Messick and Jackson seemed to have noticed—at least they said nothing about it—what would happen if the two axes of the alleged factors were rotated. Naturally, this is all in statistical fun since there is not much sense to the original factor analysis of one correlation coefficient. Two, three, or four rotations are equally possible and logical, depending on what you want the data to say. I want it to say just the opposite to Bass. So, if we rotate the axes 59 degrees clockwise using his data in (1), (the matrix on the left is Bass's original, unrotated; the one on the right is the rotated one) we get:

	Unrotated			Rotated	
	I Acquies.	II Author.		I′ Acquies.	II′ Author.
F	.77	.47	F	.01	.90
G	.53	−.47	G	.68	.21
				35%	65%

Now, instead of the statement by Bass that "... 59 per cent of the variance of the F scale is accounted for by Factor I, Acquiescence, while only the remaining 22 per cent of the reliable variance of the F scale concerns Authoritarianism as found in Factor II, "we can say that 35 per cent of the variance is associated with Acquiescence and 65 per cent with Authoritarianism!

Now suppose we play another game and rotate Messick and Jackson's correction of Bass's little table. Rotating these loadings 31 degrees counterclockwise (which corresponds to the rotation of Bass's original table given above), we get:

	Unrotated			Rotated	
	I Author.	II Acquies.		I′ Author.	II′ Acquies.
F	.77	.47	F	.90	.01
G	−.53	.47	G	−.21	.68
				65%	35%

With this "solution," we of course have the same interpretation as immediately above it, i.e., 65 per cent of the variance is associated with Authoritarianism and 35 per cent with Acquiescence, and approximately Messick and Jackson's interpretation of the unrotated matrix.

But notice what is common to both rotated matrices. The F scale is loaded only on one factor, presumably Authoritarianism. It has no loading at all on the other factor, presumably Acquiescence. The G scale, on the other hand, is loaded substantially on what is perhaps Acquiescence and negatively if only slightly loaded on Authoritarianism. This seems to land us almost back in the Adorno *et al.* lap! The analysis of one coefficient of correlation can get very complicated.

REFERENCES

1. Bass, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
2. Messick, S., & Jackson, D. N. Authoritarianism or acquiescence in Bass's data. *J. abnorm. soc. Psychol.*, 1957, **54**, 424–426.

Received September 11, 1957.

IS THE F SCALE IRREVERSIBLE?[1]

RICHARD CHRISTIE
Columbia University

JOAN HAVEL
Cornell Medical School

AND BERNARD SEIDENBERG
New York University

One of the most frequently used instruments in the area of personality and attitude measurement is the F scale developed by the authors of *The Authoritarian Personality* (1). It has not always been used judiciously, but many data have been collected in conjunction with it in hundreds of samples. It has been found to be correlated (significantly) with attitude measures varying from Arab nationalism (24) to Xenophobia (6) (not to neglect Xenophilia! [22]) and with other variables as diverse as Necker cube reversals (15) and amount of out-group aggression in dreams (19).

The F scale was designed as a covert measure at the personality level of incipient fascistic tendencies, although it is more commonly interpreted as a measure of authoritarianism.

A feature of the F scale that has come under recent critical scrutiny is the unilateral wording of its items. These are so phrased that agreement with them is scored as indicative of potential fascism, disagreement as nonfascistic. Most of these criticisms stem from Cronbach's summary of earlier work on response set and his indication of its importance in answering test or scale items (10, 11). Some subjects, it appears, have general tendencies to accept or reject items. Although this susceptibility to respond in a set way may be indicative of a broader factor of acquiescence, it need not manifest itself uniformly. According to Cronbach, response set is more likely to occur when a given item is vague or otherwise lacking in specificity or is in an area with which the individual is unfamiliar. Such considerations are pertinent to many of the items in the F scale which are characterized by nonspecificity of referent and do not make sense when analyzed logically.

RESPONSE SET AND THE F SCALE: SOME QUESTIONS

Although the authors of *The Authoritarian Personality* gave more detail about their research procedures than is customary, they failed to anticipate all the critical questions that might be raised and there are explanatory lacunae. One of these is the failure to explain why F-scale items were unilaterally worded. This is in contrast to an earlier discussion of the explicit rationale for a similar wording of the E (Ethnocentrism)-scale items since it was recognized that this violated recommendations of experts on scale construction (1, p. 59).

Three of the 38 items in the original Form 78 F scale were worded in such a fashion that rejection would imply fascistic tendencies (Nos. 12, 20, and 28: 1, Table 1 (VII), pp. 226–227). Two of them discriminated negatively (Nos. 12 and 20) and ranked 37th and 38th in rank discriminatory power in the item analysis (1, Table 3 (VII), p. 245). These two items were believed to be poor because they were "unclear or ambiguous" (1, p. 247) and were dropped from further revisions of the F scale. The third reversal (No. 28) had poor discriminatory power and was not included in subsequent revisions of the F scale. It does emerge worded in the *opposite* direction, i.e., in conformity with the rest of the F-scale items, in Form 60 A of the F scale.

The difficulties in constructing satisfactory item reversals of the F scale that are hinted at in this initial attempt are highlighted by four recent studies in which F-scale items were reversed. Correlations between the scales composed of reversed items and the original F scale (or an abridgment) indicate little success in constructing adequate reversed items. If scales composed of these reversed items measured "authoritarianism" defined as the rejection of democratically worded items it would be expected that the correlation between them and the original F scale, which was based upon agreement with authoritarian or nondemocratic statements, would approach +1.00—if subjects are ideologically consistent and the reversals are adequate. This is clearly not the case in these four studies. Bass found a correlation of but +.20 (2), Chapman and Campbell ones of +.17 and −.01 in two samples (7), Jackson, Messick, and Solley one of −.35 (14), and Leavitt, Hax, and Roche ones varying from −.66 to +.42 in five small samples (17).[2]

---

[1] This study would not have been possible without the generous aid of those cited below. Donald T. Campbell, N. L. Gage, and Douglas N. Jackson made prepublication copies of their articles available and kindly furnished further details of their research. Tentative reversals of F-scale items were critically scrutinized by Robert Holt, Marie Jahoda, Robert Lee, and Gudman Smith and their advice was in large part accepted. Angus Campbell, Richard I. Evans, Warren Miller, Lester W. Milbrath, A. Robert Rollin, and Peggy Cook played a crucial role in recruiting subjects and their aid cannot be fully acknowledged in a footnote. Finally, the services of Stanley Budnitzky, Daniel Clastner, Ethelreda Seidenberg, and Orlando Sepulveda in scoring questionnaires and performing computations are gratefully acknowledged.

Four independent studies leading to findings of negligible if not negative correlations between F scale and scales composed of reversals of its items must be taken seriously. The lack of high positive correlations was attributed to the influence of response set by all four groups of investigators. Bass even went so far as to assign three-fourths of the "reliable" variance and 59 per cent of the total variance to response set (2, p. 622). This conclusion has been challenged by Messick and Jackson who pointed out that Bass misidentified his factors when they were reflected and that a factor analysis was not justified given the nature of Bass' data (20). Although Bass acknowledged the validity of this criticism he still argued that acquiescent response set is an important component in results obtained with the F scale (4), as did his two critics.

Such an evaluation is at least partially consistent with other data. Cohn (6) was one of the first to consider seriously the possibility that the F scale was in part a measure of acquiescent tendencies. He found a correlation of +.41 between a "plus" scale of diverse items from the MMPI and an unspecified version of the F scale. Bass mentions finding only positive correlations among 13 scales of 20 aphorisms each although the content areas covered by the scales varied (2, p. 617). Items from these scales appear to be used in a scale of "social acquiescence" which is reported to correlate +.49, +.48, and +.16 respectively with the F scale in three samples (3, p. 297). Chapman and Campbell report obtaining a "rather pure" measure of response set by scoring 74 questionnaire items that were balanced in positively and negatively worded direction for a general tendency to agree. This measure was correlated with the F scale to the extent of +.43 (7, p. 131). Leavitt, Hax, and Roche cite an unpublished study by Mullery in which correlations between the F scale and acceptance of contradictory maxims of +.46, +.41, and +.38 respectively were found in three small samples (17, p. 215). A subsequent study by Carey, Rogow, and Farrell (5) on the relationship between agreement with positive and negative aphorisms and the F scale indicated that in three of four samples, agreement with both sets of aphorisms was significantly correlated with F-scale scores (positive aphorisms were defined as affirming Christian, humanist, or democratic values; negative aphorisms were those judged to deny these ideals). Webster found that a measure of response set correlated +.26 with the F scale (26, p. 31) among Vassar students.

These findings are congruent with those of Newcomb (21) in the Bennington study. He had heard Thorndike speak of the "tendency to agree" and although he was not familiar with any studies analyzing what Cronbach was later to call response set he analyzed the P.E.P. (political-economic progressivism) scale in terms of this "tendency to agree" and found that the split-half reliability on the P.E.P. scale among seniors in the class of 1939 was +.68 when scored by subtracting the number of disagreements from the number of agreements (21, p. 194). The correlation between tendency to agree as Freshmen and Seniors was +.33 (21, p. 195). These and other findings reported by Newcomb are of particular interest in the present context since the P.E.P. scale was a modification of Stagner's pioneering scale of fascist attitudes (21, p. 20).

---

[2] The sign of the correlations in these studies has been standardized in this discussion so that a positive correlation indicates consistency of content and a negative one a lack of relationship between the F scale and a reversal scale scored so that rejection of the items is scored as authoritarian.

Thorndike had actually conducted research on the tendency to agree during the 1930's. This material has not been published but an abstract of a paper given by Lorge on the topic at a regional psychological convention exists (18). Cronbach also summarizes additional material on reliabilities of scales in these studies (11). Briefly, split-half reliabilities for given response categories were high (.88 and .95), and an unspecified positive correlation was found between the number of "yes" responses on the Bernreuter and the number of statements checked on a Thurstone scale.

The preceding studies are similar in that the measure of response set is based upon statements that may vary in the extent to which they are ambiguous to respondents. In other words, the presence of response set is confounded to a certain degree with the item content.

Gage, Leavitt, and Stone have used a different technique to measure response set. They used true-false information items in a test pitched at such a level of difficulty that their respondents could not achieve better than a chance score. The split-half reliability of an acquiescent set, i.e., the number of "true" responses, was +.68. (This is identical with that found by Newcomb on the P.E.P. scale and suggests hypotheses about the nature of the political and economic attitudes of his Bennington sample.) A similar scale constructed by Gage *et al.* had items at a 75 per cent level of difficulty and the reliability was but +.09, a result indicating that "... *difficult and ambiguous items are required to elicit the acquiescent set*. . . ." (12, p. 98, italics added).

Gage *et al.* also administered, among other measures, Cohn's "plus" scale and the F scale. The latter correlated +.44, a finding similar to Cohn's. The correlation between their Information-True (items at chance level) test and the F scale was +.35. This strikes us as the most clearcut indication of the relationship between the F scale and a response set measure that is free from variance attributable to the attitudinal content of the items. Roughly an eighth of the variance in response to the F scale would then be attributable to response set—at least among studies conducted on the usual samples of American college students (the lower correlation obtained by Webster is believed to be attributable to the fact that his sample was probably more intellectually sophisticated than members of the other samples reviewed).

Although the weight of evidence indicates that responses to the F scale are affected by response set, it does not support the notion that response set is the primary determinant. The lack of consistent positive correlations between the F scale and reversed F scales in the four studies cited therefore *cannot* be attributed primarily to a tendency to acquiesce.

There are other reasons for questioning an attempt to equate responses to the F scale and acquiescent response set (2, 4). The F scale correlates highly with many measures which are free from this artifact, e.g., the analysis of incomplete sentences reported in *The Authoritorian Personality*. (One of the oversights connected with much of the justifiable methodological criticism of that research is that the methodologically more sophisticated work such as this has been largely ignored.) Another qualification should be apparent to anyone who has ever scored the F scale and observed how remarkably inconsistent many people can be in their responses—can a response set underlie such variability? Finally, most of us who have ventured outside the ivied halls have known individuals who emphatically endorse statements such as those embodied in the F scale. Yet they are not only reluctant to acquiesce to statements reflecting approval of any liberal political activity of the past quarter century but emphatically affirm that anyone making such statements is a fool if not a traitor. Among those who represent the quintessence of the F syndrome, placid agreement to explicit liberal statements is rare in our experience.

The preceding considerations indicate that acquiescent response set is one component of scores on the F scale but that its importance has been over-emphasized in the four studies using reversed items. A scrutiny of the data suggests two possible reasons. One is that the reversed items which were used have deficiencies as meaningful psychological opposites to original F-scale items. A second reason appears to lie in the distributions of F-scale scores in the samples studied; these are not congruent with the supposition that the high scoring subjects are acquiescent. These points are worthy of further elaboration.

Problems in Reversing F-scale Items

The following considerations seem important to us in the construction of item reversals which should be accepted by those who reject F-scale items and rejected by those who accept them:

1. *Logical opposition to the original items.* This point is by far the simplest and the studies under consideration are not subject to criticism on this basis. We do not, however, consider logical opposition in itself a sufficient condition for meaningful reversals. This point may be illustrated by one of the few items which was reversed in all four studies—"familiarity breeds contempt." The reversals of this laconic aphorism were:

> Familiarity does not breed contempt. [2]
>
> Familiarity breeds respect more often than it does contempt. [7]
>
> We are bound to admire and respect a person if we get to know him well. [13]
>
> The more you get to know people the more you get to like them. [17]

These statements are all *logically* incompatible with the original item. Yet they differ in wording and, we would argue, in their *psychological* incompatibility with the original statement and one another.

2. *Avoidance of extremity of wording.* The statements in the F scale are frequently couched in all or none terms: "everybody," "the most important," "never do anything," "no sane, normal, decent person," "no weakness or difficulty," etc. Since this form itself has been hypothesized as appealing to those high on the F scale and who are presumed to be intolerant of ambiguity it would appear appropriate to couch reversals in probabilistic form by using such qualifications as "some," "many," "one of the most important," etc.

In varying degrees, items used in the four studies failed to meet this criterion, e.g.:

> *No* people are born with an urge to jump from high places. [2]
>
> It is *completely absurd* to believe that some people may be born with an urge to jump from high places. [7]
>
> It is *known with complete certainty* that the urge to jump from high places is learned, not inborn. [13]
>
> *All* the mysteries surrounding our lives will sooner or later be cleared up through the progress of science. [17][3]

The theory underlying the F scale suggests that a truly nonauthoritarian individual would be taken aback by the unqualified wording of such statements. Agreement with the underlying premise that suicidal tendencies are most plausibly explained by the vicissitudes of an individual's life circumstances rather than heredity might be masked by a nonauthoritarian disinclination to agree with a dogmatic statement.

3. *Psychological opposition to original items.* Many of the reversals in these studies may be nondiscriminating because they are irrelevant to the *psychological* significance of the originals. In some instances we would predict that the reversals would be accepted by *both* high and low scorers on the F scale; in other cases that the reversals would not be acceptable to either group. Our judgments in these cases are largely subjective since they are based upon our interpretation of the theory underlying *The Authoritarian Personality* and familiarity with subsequent studies. Since it is difficult to formulate general rules in such a situation, we shall cite a few of the many statements that are relevant.

Consider two reversals of the statement: "Nowadays when so many kinds of people move around and mix together so much a person has to protect himself especially carefully against catching infection or disease from them":

> Nowadays, since democracy demands that people of widely different backgrounds and station mix together, a person should not be finicky about catching disease from any of them. [13]
>
> You need to get out and rub elbows with all kinds of people to get new ideas and broaden your understanding of life. [17]

The first reversal should presumably be accepted by low scorers since they, theoretically at least, should not be over-anxious about "contamination" by others. High scorers might also be expected to accept the reversal since it is loaded in the direction of a conforming response; rejection of the statement implies rejection of democracy, and, despite their implicitly anti-democratic ideology, their need for conformity is usually associated with the acceptance of the *symbols* of the sanctioned democratic ethos.

The second reversal, it is suggested, misses the psychological point of the original statement which was intended to arouse anxiety about catching diseases from unwashed outgroups. Instead, a folksy democratic statement was used. The basic concept is unlikely to be rejected by low scorers although they might wonder about the "need." Rejection of the item by high scorers would imply that they were intraceptive, which runs counter to theory. For these reasons, the item should theoretically be accepted by both high and low scorers and thus be nondiscriminating.

---

[3] Leavitt, Hax, and Roche did not reverse the "urge to jump" item. The italics are ours.

These considerations suggest that one of the reasons why the scales composed of reversed items fail to correlate with the F scale lies in the nature of many of the items of which they are composed. Many strike us as being subject to rejection by low scorers because of their extremity of wording, irrationality, or lack of compatibility with the psychological significance with items in the F scale. Other reversed items are believed to be nondiscriminating because they might reasonably be expected to be accepted by both high and low scorers. There are enough items subject to these strictures in the studies under review to afford a plausible explanation for the findings reported without the necessity of invoking response set as *the* crucial variable. These considerations are believed to be especially crucial for the studies of Bass, who explicitly attempted to construct *logical* rather than *psychological* reversals, and of Jackson and Messick whose reversals were in the *style* of the original items.

Acquiescent Response Set and Sampling

Three of the four samples under review used college samples, apparently largely composed of students at the beginning of their college training. The majority of respondents in the fourth (17) appear to have been highly educated. This point is of methodological interest because college samples do not tend to make *absolutely* high scores on the F scale. More precisely, if the range of item mean scores is defined as ranging from 1.0 (strong rejection of every item in the scale) to 7.0 (strong acceptance of every item), it is found that most college samples reported in the literature have a mean score below the theoretical neutral point of 4.0.

This distribution of scores is important in evaluating the purported relevance of response set to the F scale. Those in the lowest quarter of the typical distribution usually have an item mean score of well below 3.0 which indicates that they must fairly consistently disagree "slightly" or "somewhat" with the items with an occasional agreement or "strong" disagreement. Those who score in the highest quarter typically have mean scores in the 4.0 or 5.0 range with a few scoring slightly higher.

The importance of such distributions for the equation of high scores on the F scale with acquiescence is due to the fact that "neutral" responses are not customarily allowed. Any scores between 3.0 and 5.0 therefore have to be obtained by balancing off agreement and disagreement with specific items. The typical college student "high" authoritarian *cannot*, by virtue of the scale format and scoring, *be consistently accepting items*. This fact suggests that if these studies are based upon typical college samples, the lack of positive correlations reported cannot be attributable to acquiescence response set on the part of the high scorers in these samples since the latter are inconsistent in their responses. (It is hypothetically possible that some of the reported correlations are due to "negativism" on the part of low scorers who reject both the original F-scale items *and* the reversals!)

It is impossible to delineate fully the possible effects of the range of responses on the F scale in the samples used in these four studies. Bass (2) reports an item mean of 2.83 which is equivalent to one of 3.83 in the conventional scoring of the F scale (a constant of three rather than four was added to raw scores). He reports an *SD* of 2.00 (2, Table 2, p. 620). This means, if correctly reported, that his sample was necessarily uniquely bimodal on the F scale since it would be impossible to have a normal distribution on a scale with a six-point range and a standard deviation of a full two points. In view of this and other errors in these data, no reliable estimates can be made about the characteristics of the sample.

Chapman and Campbell do not report the mean or distribution of F scale responses for their two samples. They used a five-point response scale with an undecided category so that the preceding comments about the lack of a neutral category do not apply to their data. An item mean computed from unpublished data indicates that the item mean for their second sample was 2.90, just below the theoretical neutral point of 3.00. Since these data were in the form of responses to items, it was not possible to compute an *SD* based upon *S*s. It is therefore impossible to determine the extent to which high scoring *S*s were consistent in their acceptance of F-scale items.[4]

Jackson, Messick, and Solley report an item mean of 3.10 with an *SD* of .67 on 42 *S*s. Their scoring procedure differed slightly from the conventional one and yields a mean equivalent to 3.77. If their distribution of scores was normal, approximately one or two *S*s would have scored above 5.0. Leavitt, Hax, and Roche studied two small samples of foremen, (n's = 6 and 13), two of executives (n's = 34 and 52), and one of Unitarians (n = 28). A composite plot of their data (17, Fig. 1, p. 219) indicates that low scoring *S*s greatly outnumbered high scoring ones. As nearly as we can determine by inspection, eight *S*s had an item mean agreement varying between "slight" and "somewhat" and a single *S* had an item mean above 6.0. The relatively high scorers in these samples thus fell largely within or slightly above the neutral range.

Additional Comments

It should be emphasized that the preceding comments about these studies are based upon an advantage that the original authors did not have, i.e., the opportunity to be familiar with one another's work. The seemingly paradoxical finding in independent studies that F-scale reversals did not correlate appropriately with the F scale raised questions which led to analysis of possible reasons. Chapman and Campbell have an insightful discussion of the problems of reversing items and because of the difficulties did not try to reverse more than ten items in their initial attempt (7, p. 129). Since they were the only investigators to report an item analysis, a study of their items which did successfully reverse gave us clues as to the possible failures of their other items and those of the other investigators. Further hunches were suggested by Bass' analysis of reversed items as to the degree of judged logical difference from the originals, the strikingly negative correlation found by Jackson *et al.* with reversals that were stylistically similar to F-scale items, and the different correlations among samples reported by Leavitt *et al.*

An Exploratory Investigation of Reversed F-Scale Items

The present study was undertaken to explore the implications of our examination of earlier studies. Clarification of the problem of the role of response set in the F scale required, we believed: (*a*) reversed items which were constructed according to the criteria previously specified and (*b*) the testing of these items on samples believed to contain a wide range of scorers on the F scale.

The Item Pool

It was decided that the larger the number of reversals attempted, the greater the probability of empirically finding ones which were correlated with the F scale. In two studies (7, 17), the researchers did not test reversals of all the items in the F scale. In three of the studies (2, 13, 17), F-scale items were reversed although they had been explicitly mentioned in *The Authoritarian Personality* as being poor discriminators (1, Nos. 22 and 44, pp. 254–257). We attempted to reverse every F-scale item which was acceptable in its original form.

The final version of the F scale described in *The Authoritarian Personality* contained 29 items. We attempted to reverse the 27 which were found to be discriminating. One, "wars and social troubles may someday be ended by an earthquake or flood that will destroy the whole world," baffled us in our attempts to construct a meaningful reversal. Three items which were in the subsequent Form 60 A were added to the pool. A fourth additional item, taken from the earlier Form 60, was also added. It had discriminated between high and low scorers (1, Table 6, (VII), item 41, p. 258) but, for unstated reasons, was not included in subsequent revisions of the F scale.

---

[4] The item mean is higher than generally found in schools of Northwestern's caliber. Since Chapman and Campbell did not randomly sample F-scale items and used but 16 of them, the possibility of bias in the items was considered. The 14 items taken from the Form 40 of the F scale had an item mean (our computations) identical with the other items in this form in the original samples (based upon 1, Table 9 (VII), p. 260). Their other two items were taken from the Form 60 and had an item mean of 4.46 in contrast to that of 3.71 on the whole scale (1, Table 6 (VII), p. 253). Since there is reason to believe that the relative "pull" of F-scale items is fairly consistent in college samples, a slight contribution to the unexpectedly high item mean in this sample may be attributed to the popularity of these two items. The effect of allowing a "neutral" option upon item mean scores is unknown.

The Item Reversals

After discussion of the problems of item reversal along the lines indicated above, the authors individually attempted to write one or more reversals of each item in the pool. These reversals were then discussed and one of them, or more often a modification of one, was accepted for provisional inclusion in the reversed scale. The reversals were then circulated among colleagues familiar with the F scale; their critical comments were solicited and, in most cases, accepted.

The Subscales

Two questionnaires (Forms A and B) were prepared. Each was composed of 15 of the 30 original F-scale items and reversals of the other 15 since it was believed that the presence of the original and reversed form of the same item in the same questionnaire would lead to possible artifacts in response. The assignment of items to the two forms was based upon the use of a table of random numbers and the serial numbering of the items within the questionnaire was also random. There was no a priori reason for assigning items since earlier research had indicated that individual items in the F scale fell into different empirical clusters in different samples (8). Each scale was a mirror image of the other, i.e., if a given item appeared in its original wording on Form A its reversal was in the same serial position on Form B. The item reversals used, their allocation to Form A or B, their position within the subscales, and the source of the original item are given in Table 1.

Scoring

The customary procedure in scoring the F scale was followed, respondents being given an option of agreeing strongly, somewhat, or slightly or disagreeing slightly, somewhat, or strongly. No neutral or undecided option was allowed, but omission of an answer was scored as neutral. Both scales were scored so that high scores reflected "authoritarianism" as indicated in Table 2.

The Samples

We were interested in obtaining samples which included *S*s who would show great variability of scores on the F scale and were especially interested in finding high scorers. Like other investigators, we were hampered by the difficulty of obtaining noncollege samples. Through the cooperation of friends, the scales were administered to their classes at two colleges that were believed to contain a higher proportion of high scorers than would be found in most colleges with which we were familiar. One of these was a private university in the Southwest, the other, a Northeastern public university. The scales were administered during the 1956 summer session when the student bodies were more heterogeneous than usual since they also contained many elementary and high school teachers who were attending summer school.

*S*s in each of the two college samples were divided in half by administering odd (Form A) and even (Form B) scales to adjacent *S*s. Results on four samples will be discussed: SW-even ($n = 75$), SW-odd ($n = 75$), NE-even ($n = 39$), and NE-odd ($n = 40$). Item mean scores and *SD*s on the 15-item F subscales were respectively: 3.81, .97; 3.39, .93; 3.26, .99; 3.35, .93. We are unable to explain why the SW-odd sample scored so surprisingly much lower than the SW-even sample since the alternate forms were distributed to adjacent students and the two SW samples do not differ significantly in proportion of males (roughly two-thirds), age, or major in college. Roughly half of the NE samples were male. Fourteen per cent of both samples were psychology majors, 25 per cent were education majors, and 19 per cent were business adminstration majors. The bulk of the remainder of the NE samples was composed of nursing students, that of the SW samples of students in engineering and liberal arts. Respondents ranged from freshmen to graduate students in all samples.

RESULTS

Item Analysis

In analyzing the reversals, scores made by the *S*s on the appropriate F subscale were divided into quarters. Two comparisons were then made of the scores on individual reversed items—the magnitude and direction of the individual item mean differences of those in

TABLE 1

REVERSALS OF F-SCALE ITEMS

Orig. No.[a]	Form	No.	Reversed Statement
31	B	1.	There are times when it is necessary to probe into even the most personal and private matters.
37	A	2.	It would be a good thing if people spent more time thinking and talking about ideas just for the fun of it.
21	A	3.	If it weren't for the rebellious ideas of youth there would be less progress in the world.
9	B	4.	When a person has a problem or worry, it is best to face it and try to think it through, even if it is so upsetting that it keeps him from concentrating on other things.
19	B	5.	Insults to our honor are not always important enough to bother about.
2	B	6.	There are many difficulties a person cannot overcome no matter how much will power he has.
39	B	7.	It's nobody's business if someone is a homosexual as long as he doesn't harm other people.
8	A	8.	It's all right for people to raise questions about even the most sacred matters.
29	B	9.	It is highly unlikely that astrology will ever be able to explain anything.
13	A	10.	In the long run it is better for our country if young people are allowed a great deal of personal freedom and are not strictly disciplined.
18	B	11.	Even though people of all sorts mix together nowadays, you don't have to worry very much about catching an infection or disease.
34	B	12.	Even if there were no immoral, crooked, or feeble-minded people there would still be many serious social problems.
16	A	13.	An urge to jump from high places is probably the result of unhappy personal experiences rather than something inborn.
26	A	14.	It doesn't make much sense to divide people into groups like the weak and the strong; too many people are strong in some ways and weak in others.
4	A	15.	The findings of science may some day show that many of our most cherished beliefs are wrong.
41	A	16.	People ought to pay more attention to new ideas, even if they seem to go against the American way of life.
43	B	17.	You may dislike a person very much, but the chances are that if you get to know him well you'll have more respect for him.
1	A	18.	One of the most important things children should learn is when to disobey authorities.
59	A	19.	Most people can be trusted.
53	B	20.	You'll do better work in the long run if you do things your own way rather than by following tried and true procedures.
23	B	21.	It is more important for this country to have a just legal and political system than a series of trustworthy leaders, however courageous, tireless, and devoted they might be.
25	A	22.	Sex crimes, such as rape and attacks on children, are signs of mental illness; such people belong in hospitals rather than in prison.
38	A	23.	Many people have too great a fear of plots hatched in secret by politicians.
42	A	24.	It's only natural for people to sometimes have thoughts about hurting a close friend or relative.
41	B	25.	The artist and professor are probably more important to society than the businessman or manufacturer.
27	A	26.	Most honest people admit to themselves that they have sometimes hated their parents.
6	B	27.	Human nature doesn't make war inevitable; man may some day establish a peaceful world.
35	B	28.	In spite of what you read about the wild sex life of people in important places, the real story is about the same in any group of people.
12	A	29.	People should be willing to overlook failures in the manners and unpleasant personal habits in other people.
58	B	30.	Books and movies ought to give a more realistic picture of life even if they show that evil sometimes triumphs over good.

[a] All items are taken from 1, Table 7 (VII), pp. 255–257 except for No. 16 which was found in Table 4 (VII), p. 250 and Nos. 19, 20, and 30, the originals of which are in 8, Table 3, p. 464.

the upper and lower quarters on the F subscales and the pattern of differences in all four quarters.

Of the 30 reversals, 22 discriminated in the appropriate direction between high and low quarters on the F scale in both samples responding to them, four in one sample but not the other (Nos. 7, 11, 14, and 17), and four discriminated negatively in both samples responding to them (Nos. 1, 20, 21, and 23). This differs from chance expectancy at better than the .01 level as tested by chi square. The attempt to find workable reversals is thus viewed as moderately successful.

Individual F-scale items were similarly analyzed against extreme quarters on the total scale. All discriminated in the appropriate direction but there was great variability in discriminatory power and mean level of acceptance.

An Examination of Response Set

Although these findings indicate that F scale items are, at least within limits, reversible, they do not clarify the more basic problem as to how response set is linked with the F scale. Given the fact that the F scale has been found to be correlated with a host of measures that do not involve response set, that its construct validity is tenuous, and that response set plays some role—although not the dominant one in the acceptance or rejection of its items—how best to proceed?

Since we had selected our reversed items against the criterion of scores on F subscales we decided to examine further the most successful five in each form. This procedure eliminated the negatively discriminating items and those of marginal discrimination and left only those found to be systematically related to response to the F scale.

We also pared down the number of F-scale items for more intensive scrutiny to five in each subscale. These were the ones with the highest discriminatory power in the item analysis. So doing tended to get rid of not only those items which were poor discriminators but those which were widely rejected or accepted. For example, the item mean on the statement, "People can be divided into two distinct classes: the weak and the strong," was but 1.77 in the SW-even sample and even those scoring in the high quarter tended to reject this "authoritarian" item, their item mean score being but 2.0 (disagree "somewhat"). On the other hand, the SW-odd sample tended to accept the statement, "The wild sex life of the old Greeks and Romans was tame compared to some of the goings-on in this country, even in places where people might least expect it," the item mean being 5.15. Slightly greater acceptance than rejection of this statement on the part of low scorers was found (item mean of 4.20). The ten reversed items (five from each form) used in this analysis, were Nos. 3, 5, 13, 15, 16, 18, 25, 28, and 30 in Table 1. The F-scale items were the originals of Nos. 1, 6, 8, 9, 15, 18, 20, 22, 24, and 30.

These short five-item scales had a mean part-whole correlation of .82 with the longer 15-item forms and thus may be viewed as tapping the essence of the F scale. They are also useful since they are presumably the most susceptible to response set if response set is part and parcel of the "authoritarian" syndrome. After all, items that are universally accepted or rejected cannot yield much information about differential response set among individuals.

TABLE 2

SCORING OF STRAIGHT AND REVERSED F-SCALE ITEMS

Response	Score: Straight F	Score: Reversed F
Agree Strongly	7	1
Somewhat	6	2
A Little	5	3
No answer	4	4
Disagree a Little	3	5
Somewhat	2	6
Strongly	1	7

TABLE 3

CORRELATIONS BETWEEN LONG AND SHORT FORMS OF STRAIGHT AND REVERSED F-SCALE ITEMS

Sample	Scales Being Correlated: 15 item F+ and F−	15 item F+ and 5 item F−	5 item F+ and 5 item F−
SW-odd	.35	.42	.45
SW-even	.14	.27	.23
NE-odd	.34	.27	.37
NE-even	.45	.66	.44

The correlations between these abbreviated scales and the original 15-item scales is given in Table 3. Since the items in the short scales were selected against the longer F subscales, it is not surprising that the magnitude of the correlations is similar. It is clear, however, that the correlations fall far from what would be expected if the reversals were adequate and response set did not play some role.

Scores of the 299 *S*s in the four samples were grouped and ordered in terms of their item mean response on the F scale. The 58 *S*s making scores of 5.0 and over were considered to be high scorers and their responses to the reversed items were examined. Sixteen of these individuals had item mean scores of 6.0 or higher on the short F scale. Despite this high degree of "acquiescence" with "authoritarian" items, it did not generalize to the reversed items. Their scores here all fell in the neutral range (3.2 to 4.8) with the exception of one slightly deviant case with a score of 5.0 (disagree slightly). There are many ways of balancing off responses to attain a "neutral" score and these *S*s displayed a most consistent inconsistency in doing so. They used the extreme categories in answering the reversed items; 84 per cent of their total responses were in terms of "somewhat" or "strong" acceptance *or* rejection. No one accepted or rejected more than three or less than two of the reversed items. Further, there was no observable difference in the acceptance or rejection of specific items.

It is clear that an acquiescent response set does not account for these results. These individuals who wholeheartedly and consistently endorsed the most discriminating F-scale items displayed an enthusiastically random pattern of response to the reversed items. Our interpretation is that they have a tendency to make extreme responses and are ideologically somewhat confused.

Those scoring from 5.0 to 5.8 ("slight" to almost "somewhat" agreement) on the short F scale differ in showing extreme variability on the reversals, the item mean range varying from 1.8 to 5.6. Those at the extremes of this continuum display quite different behavior.

Five of these individuals had item mean scores of over 5.0 on both the F and reversed F subscales, a higher score on the reversed scale, incidently, than any of 16 individuals with scores of 6.0 or over on the F subscales considered earlier. These are clearly ideologically consistent individuals and are moderately fascistically oriented. Their modal response pattern is that of accepting four out of five items on the F subscales and rejecting four out of five on the reversed F subscales. The small *n* does not suffice for an evaluation of the intensity of response although there is a lack of use of extreme categories (there was not a single "strong" approval of a reversed item). This fraction of the *S*s shows no evidence of response set.

Acquiescent response set is definitely manifested in a group of 13 *S*s with F subscale scores above 5.0 and reversed subscale scores of 3.0 or below. The modal pattern was acceptance of four of five items on *both* the F and reversed F subscales. The "authoritarianism" manifested by these individuals on the F scale appears nonideological in nature and a reflection of a tendency to agree with incompatible items.

The remaining 24 *S*s in this range of scorers show a mixed pattern of responses, the mean on the reversed F subscales falling between 3.2 and 4.8. Although tendencies in the response set direction may be present among some of them, it is equally plausible that they tended toward a consistency which was masked by the analysis of a relatively small number of items.

A similar analysis was made of those scoring low on the F subscale. There were 21 *S*s who disagreed "somewhat" or greater (scores of 2.0 or below). Thirteen of these had a general pattern of acceptance on the reversed items and may be considered consistent lows. Eight, however, had mean reversed scores in the neutral range and thus resemble mirror images of our extreme high scorers. They differ from the latter, however, in not predominantly utilizing extreme responses to items: 38, 40, and 23 per cent of their responses fell in the "slight," "somewhat," and "strong" agree and disagree categories. There is thus no evidence of response set among these respondents.

The moderately low scoring *S*s who scored from 2.2 to 3.0 on the F subscales can be identified as ideologically consistent or as subject to response set just as those with moderately high scores. A total of three *S*s displayed a consistent *negative* response, i.e., they tended to disagree with items on *both* subscales. Eighteen *S*s were identified as consistently low since they accepted the reversed items while rejecting the original ones. The remaining 20 *S*s in this range had scores in the neutral range on the reversed scales and are considered a mixed group.

Within these samples, at least, three general findings emerge: (*a*) there is a much higher proportion of ideological consistency among low than high scorers, (*b*) response set is more prevalent among those who moderately agree than those who agree extremely and although slightly present among those who moderately disagree is not found among extreme disagreers, and (*c*) there is a higher proportion of ideological inconsistency on the reversed items among extreme high scorers than among other groupings.

The consistency of the low scorers is congruent with the findings of Leavitt *et al.* They combined their groups and correlated the original F-scale items used and their reversed scale and found a positive correlation in the bottom third of the combined samples of +.45 (the high third on the F scale had a correlation on the two scales in the *negative* direction: −.32). Although we have not replicated their analysis, an inspection of our data indicates (as can be easily guessed from the earlier discussion of response patterns) that the low scorers are responsible for the overall positive correlations; the correlation between original and reversed items among the high scorers in these samples would not differ appreciably from zero.

The greater presence of response set among those with moderate rather than extreme attitudes is not at all surprising. This is precisely what would be expected according to Cronbach's formulation of the conditions under which response set is most likely to be found. A person sophisticated enough to reject most of the F-scale items out of hand has a well enough formulated ideology to know where he stands and is unlikely to find reversed items vague or unfamiliar. An extreme high scorer is of different persuasion but his devotion to the irrational does not indicate bland acceptance of contrary statements, at least in light of current theory and knowledge. Although the preceding point seems exceedingly obvious, it has not been prominent in recent discussions of response set. One possible value of the present findings, limited as they may be, may be the reiteration of basic considerations about the relationship between extremity and intensity of attitudes. These, in conjunction with Cronbach's observations, suggest that it is precisely those persons who do not have strong attitudes pro or con who should be most apt to display response set.

The one unexpected finding was the complete abandon with which the extreme high scorers rejected or accepted the reversed items. There is considerable discussion in *The Authoritarian Personality* about the irrationality of such individuals but no indication that it would take such a remarkable form.

FURTHER STUDIES

Although the preceding findings appeared plausible, they were descriptive and left several matters in doubt. First, how stable were they? Our procedure of concurrently analyzing items from separate samples is similar but not identical to doing an item analysis on an original sample and cross-validating on a new one. Further, our results were obtained on samples of summer school students who can be viewed as representative of neither the more usual college samples nor of the general population. Replication of the findings in other samples therefore appeared called for.

The second matter of concern was the unanticipated behavior of the extreme high scorers on the reversed items. Some confusion would not have been surprising but the failure to find a single high scorer who would firmly reject reversed items was most puzzling. Since these individuals were key members of our samples whose responses served as the basis for the item analysis, it was not clear whether this might be due to an artifact arising from the method of item selection, whether it reflected the possibly aberrant nature of our samples, or whether it was indeed a characteristic of high scorers on the F scale.

Data have therefore been collected on eight additional samples. These, in all candidness, were "targets of opportunity," i.e., they were not systematically selected to test specific hypotheses about the relationship between the straight and reversed forms of the F scale. They are samples made available informally through the kindness of friends in testing their classes or in permitting the inclusion of the scales in ongoing studies directed toward other ends.

The Scales

The reversed F scale was composed of Items 3, 9, 13, 15, 16, 18, 25, 26, 28, and 30 in Table 1. With the exception of Item 26, these are identical to those used in the subscale analysis in the exploratory study.

The F-scale items were the originals of Nos. 1, 6, 8, 10, 12, 20, 22, 23, 24, and 27 in Table 1. They differ from those used in the subscale analysis in that the originals of Items 9, 15, 18, and 30 were deleted and substitutions were made to prevent the presence of a reversal and its original in the same questionnaire.

Ordering of the items was random. The customary response options were provided. All samples received the same 20-item scale under conditions noted in the description of each.

The Samples

1. Undergraduate students in a Northeastern public university were given the 20-item scale during the first week of a course in introductory psychology. It was administered by the instructor (male) in charge of the classes with assurances about the preservation of anonymity and the use of the results for research. The respondents were young (median age of 18.7 years), 72 per cent were male, and they were predominantly lower middle-class in background. Sixty-two per cent were freshmen and 30 per cent were sophomores.

2. Students in a graduate seminar in social psychology at a major Eastern university filled out the 20-item scale and also were required as part of the course to administer it to friends and acquaintances. Protocols of individuals familiar with the F scale were discarded. The sample as a whole did not respond to the questionnaire under standardized conditions. It is unusually high in years of formal education.

3. Students in a summer school course in remedial reading at a prominent Eastern teachers college filled out the 20-item scale at the end of a class period. It was administered by a female introduced by the instructor as carrying out research. Most of the respondents were female school teachers from a broad geographic area and had a median of roughly 15 years of teaching experience. Most of them had done their undergraduate work in teachers colleges and were taking the course to maintain teaching credentials.

4. A nonrandom but fairly representative sample of registered Washington lobbyists was given three separate questionnaires to be filled out and returned after each had been interviewed about his role as a lobbyist. One of the questionnaires contained six F-scale items and four reversals. These were interspersed with 25 other items constructed by the

senior author which were largely taken from the writings of Machiavelli. A five-point response scale which included a neutral or undecided category was used. The majority of this sample held law degrees and most had previously held elective or appointive political office.[5]

5–8. These samples consisted of senior medical school students (with one junior class) in three top-ranking medical schools. Students in two which were located in the East had predominantly received their undergraduate training at the more prestigeful Eastern liberal arts colleges. Those in a Midwestern school were largely from the better Midwestern colleges.

Items from the 20-item scale were randomly distributed among 20 Machiavellian items (half of which were worded in the reversed direction) in an "opinion inventory." These and other psychological measures were part of a lengthy questionnaire devoted to experiences in medical school. In some instances the questionnaire was answered in a class setting, in other cases individuals filled it out at home.[6]

Results

The data summarized in Table 4 indicate the consistency with which positive correlations between the F and reversed F scale were found. The samples varied in education from beginning college students to those with eight years of undergraduate and professional experience in the finest of our educational institutions, from predominantly female elementary and high school teachers to lobbyists with a bad case of "Potomac Fever," some answered the questionnaires anonymously in class while others filled them out in conjunction with personal interviews, the items were sometimes presented as an individual scale and were sometimes mixed with other items and in five samples were given along with a large number of unrelated questions, and in general the customary seven-point scoring system was used, but in one case five-point response categories with a neutral option were used.

Despite all these variations that might possibly affect the nature of the responses positive correlations were found in all eight samples. When these are combined with the findings in the exploratory study it is found that in twelve of twelve samples using various combinations of these items the results are such as to indicate that response set is not important enough to produce negative correlations.

The other problem that intrigued us—that of the behavior of the extreme highs on reversed items—is partially clarified by a closer inspection of these data. The distribution of responses on the F scale was leptokurtic in the undergraduate sample, there being but eleven individuals with a mean score of 3.0 or below and five with scores of 6.0 or above. The heavy concentration of respondents in the neutral and slightly accepting range casts some light on the very low correlation between the scales in this sample since the generalizations made earlier about the relationship between mean acceptance of F-scale items and acquiescence were confirmed. However, of the five high scorers, two displayed acquiescent tendencies, two behaved similarly to the extreme highs discussed earlier, and one was consistent.

Further evidence that the extreme high was not necessarily inconsistent came most strongly from the lobbyist sample where there was very little evidence of acquiescence. Indeed, we found one "perfect" high scorer (within the limits of the items used) who strongly accepted all six F-scale items and strongly rejected all four reversals. Unfortunately, there were not enough extreme high scorers among these samples to serve as a sound basis for generalizations.

The wide range of reliabilities on both the original and reversed F scales and the varying correlations between them posed problems in interpretation. It should be added that where data from other attitude scales were available, they indicated that the amount of ideological consistency present was not unique to the F scale. The Washington lobbyists, for example, had the highest internal consistency on the Machiavelli scale of nine samples totaling some 2,000 *S*s tested to date. This finding parallels their results on the F scale. Similarly, Sample 5, composed of a particularly aberrant (statistically speaking) group of medical students, showed inconsistency similar to that on the

[5] This study of the role and personality attributes of lobbyists was carried out under the auspices of the Brookings Institution under the supervision of Lester Milbrath.

[6] The samples were studied as part of ongoing research on the Sociology of Medical Education by the Bureau of Applied Social Research at Columbia University under a grant from the Commonwealth Fund.

TABLE 4

SUMMARY RESULTS ON F AND REVERSED F IN EIGHT SAMPLES

Sample	N	F Scale			Reversed F Scale			Correlation between F and Reversed F Scales[a]
		$\overline{M}$	$\overline{SD}$	Rel.	$\overline{M}$	$\overline{SD}$	Rel.	
1. Northeastern undergraduates	152	4.54	.84	.52	3.73	.78	.43	.10
2. Graduate students	57	2.74	.95	.78	2.72	.93	.57	.58
3. Lobbyists	86	3.05[b]	.59	.54[c]	3.03[b]	.81	.41[d]	.55
4. Remedial reading teachers	61	3.43	1.05	.68	3.20	.94	.60	.29
5. Midwestern medical students (seniors)	72	3.50	.69	.34	3.35	.55	.18	.45
6. Midwestern medical students (juniors)	80	3.55	.88	.66	3.47	.70	.26	.23
7. Eastern medical students[1] (seniors)	73	3.38	.92	.75	3.35	.91	.72	.38
8. Eastern medical students[2] (seniors)	121	3.76	.86	.58	3.17	.68	.42	.48

[a] All correlations were corrected for attenuation.

[b] A five point scale was used in this sample. These figures are comparable to slightly over 4.0 in other samples.

[c] Based on only six items.

[d] Based on only four items.

F scale when responding to Machiavellian items but were very consistent on a measure of manifest anxiety. Such findings suggest that samples vary widely in the amount of acquiescence that they will display in response to items of an ideological nature.

The investigation of differences from sample to sample demands a more refined technique of measuring response set than that used in the preceding examples. Before pursuing the substantive problem, it is necessary to indicate the method used.

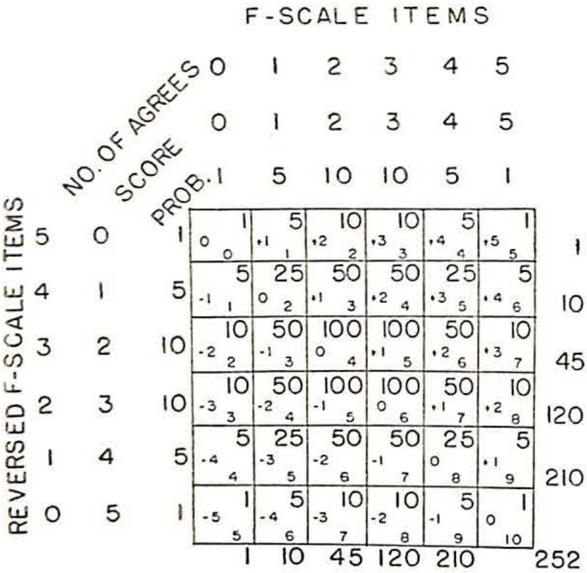
A METHOD OF MEASURING ACQUIESCENCE

The diagnosis of acquiescence used with the four exploratory samples was crude and obscured the interaction of content and response style. The correlations and reliabilities indicated in Table 4 suggest widely varying probabilities of response set in different samples. These respondents were subjected to different numbers of items and in some cases different choices of response categories. In considering the problem of comparability from sample to sample, it was decided that it was simplest to use the number of statements agreed to on the F and reversed F scales as the basic data. This convention permitted the use of the binomial expansion and the ascertainment of the probability of a given pattern of response.

Although the technique is basically simple, we are unfamiliar with examples of its application and are illustrating its use with two five-item scales.

There are six possible degrees of agreement for each five-item scale. Cross-multiplication of the binomial frequency for each response category on each scale indicates the chance probability in 1024 times of the frequency of a given cell in the matrix, e.g., the probability of a given individual agreeing with all five items in each scale is one in 1024. The expected frequency of any given pattern of response is indicated in Fig. 1.

The second information in the matrix is the discrepancy scores between the two scales. If the two scales are measuring essentially the same thing—and the means on two sophisticated samples (graduate students and lobbyists) as indicated in Table 4 suggest that this is approximated—then an individual who agrees with all five F-scale items should tend to disagree with all five reversed items. A simple way to check for consistency is to reverse the number of agreements on the reversed scale so that agreement with all five is scored 0 and disagreement with all is scored 5. Subtraction of the reversed scale score from the F-scale score gives a measure of content consistency. A perfectly consistent individual should have identical scores on both scales. A person who is indiscriminately accepting will have a discrepancy score of +5. If he is recalcitrantly negativistic he will have a discrepancy score of −5.



REVERSED F-SCALE ITEMS			F-SCALE ITEMS						
NO. OF AGREES			0	1	2	3	4	5	
SCORE			0	1	2	3	4	5	
PROB.			1	5	10	10	5	1	
5	0	1	1 0 0	5 +1 1	10 +2 2	10 +3 3	5 +4 4	1 +5 5	1
4	1	5	5 -1 1	25 0 2	50 +1 3	50 +2 4	25 +3 5	5 +4 6	10
3	2	10	10 -2 2	50 -1 3	100 0 4	100 +1 5	50 +2 6	10 +3 7	45
2	3	10	10 -3 3	50 -2 4	100 -1 5	100 0 6	50 +1 7	10 +2 8	120
1	4	5	5 -4 4	25 -3 5	50 -2 6	50 -1 7	25 0 8	5 +1 9	210
0	5	1	1 -5 5	5 -4 6	10 -3 7	10 -2 8	5 -1 9	1 0 10	
			1	10	45	120	210		252

FIG. 1. MATRIX FOR DETERMINING THE PROBABILITY OF A GIVEN PATTERN OF RESPONSE ON 5 ITEMS F AND REVERSED F SCALE

The number in the upper right hand corner of each cell indicates the chance expectancy in 1024 times of a given pattern. The upper small number indicates the discrepancy score; the lower one the total F score on the combined scales. The numbers at the right and bottom are the summed frequencies for the cells with identical discrepancy scores along the diagonals of the matrix.

The next step is to determine the probability of occurrence of a given discrepancy score. Summing along the diagonals composed of cells with the same discrepancy scores gives the appropriate frequencies. It will be noted that this distribution is the binomial expansion of 2^{10}. The probability of occurrence of a discrepancy score of +3 or greater is .055 ([1 + 10 + 45]/1024). This is as close to the more customary .05 level as can be obtained and is sufficient for present purposes. Since the matrix is symmetric, discrepancy scores of −3 or greater indicate a tendency to disagree significant at the same level.

A combination of scores on both scales can also be used to obtain a measure of overall F by scoring along the orthogonal diagonal. An individual who rejects all five of the original items and accepts all five reversals is clearly low on the dimension and is given a score of 0 as indicated in Fig. 1. His opposite number obtains a score of 10. Scores of 2 or less and 8 or more differ significantly from chance expectancy.

Acquiescence in the College Trained Population

A clearer picture of the operation of acquiescence response set may be had by introducing data collected on one further sample. This is a subsample of the 1956 post election survey conducted by the Survey Research Center of the University of Michigan. Adult respondents selected in a nationwide probability sample were asked their education and 103 reported having at least some college training. In the course of an interview they responded to five F-scale items and five reversed items.[7] The preceding method of analysis was employed and it was found that five per cent were significantly acquiescent, 28 per cent slightly so, 23 per cent were completely consistent, 33 per cent showed slight negative tendencies, and 11 per cent had a significant negative response pattern. It is of considerable interest, in light of current equation of acquiescence with response set, that this sample displayed slightly more negativism than acceptance in responding to items. The amount of observed response set does not, however, differ from chance expectancy.

This finding takes on possible significance since graduate students and Washington lobbyists also displayed a negligible amount of response set. These are samples which are presumably the least confused ideologically. The sample which is most distinctive is that composed of Northeastern undergraduates. A technique similar to that indicated in Fig. 1 was employed for the two ten-item scales they answered. The probability of having a discrepancy score of ±4 or greater is .059.

Thirty-five or 23 per cent of these students showed a significant amount of acquiescence and five (three per cent) showed significant negative tendencies. Although the definition of significance level is slightly different for these undergraduates and the nationwide sample, the difference is quite striking. Why should undergraduates be so much more acquiescent than a nationwide sample of adults exposed to college (only two-fifths of whom were graduates)? A precise answer would require a detailed investigation of factors influencing changes over time. It should be noted, however, that such a finding is congruent with a variety of data on attitude changes in the first years of college. Jacob has summarized a number of studies dealing with changes on the F scale and other measures such as favorableness toward civil liberties and other attitudes indicative of liberalism (16). The general finding is that students are most conservative when they enter college. Exposure to the more liberal norms of college usually leads to high scorers dropping out of school, changing in a more liberal direction, or adamantly holding to their original beliefs (Newcomb has furnished graphic illustrations of the last two reactions [21]). The weight of the evidence indicates that this conflict of values is most severe during the initial year or so of college.

---

[7] The relationship between education and response set in the entire sample is currently being analyzed and will be reported in a separate publication. The items were slightly modified from the wording indicated in Table 1 for use with respondents of varying verbal sophistication. Details of the sampling procedures and the analysis of the interview material will be found in the forthcoming book on the 1956 presidential election by Angus Campbell and Warren Miller. This study was carried out by the Survey Research Center with a grant from the Rockefeller Foundation.

Our tentative interpretation is that many among these undergraduates—who were in the very early days of college—were in a stage of value transition and possibly conflict. They still tended to accept F-scale items which are more widely accepted in noncollege circles such as those from which they had just come. At the same time, the reversed items were presumably more in keeping with the value climate of the university to which they were adjusting. Many of the students tended to accept both sets of items. This "acquiescent response set" does not strike us as a mere mechanical way of responding; it probably reflects another facet of Cronbach's discussion of the ambiguity of an item. In this case, items reflecting contradictory views of the world are accepted because the student has not yet decided for himself which is *the* right response; both are supported by different reference groups.

DISCUSSION

Although the preceding digression is slightly speculative, it has important implications if valid. An initial question about some of the studies emphasizing response set focused upon the actual scale scores of the *S*s and the fact that "high" scorers appeared to be only relatively high and were in actuality inconsistent in their responses. Upon the basis of these later considerations, based, it is true, upon a limited sampling of undergraduates, it would appear that the majority of these studies have used subjects who were most liable to display response set of any populations with exposure to higher education, i.e., undergraduates at an early stage of training.

There is another obvious implication of the relatively high number of acquiescent respondents in the undergraduate sample. Most of the studies involving *S*s high and low on the F scale use samples drawn from undergraduate populations and most usually those in the first two years of college. This means, quite bluntly, that the high scorers are quite likely a mixed group. Some are genuine high scorers with a consistent set of authoritarian attitudes that involve rejection of reversals as well as acceptance of the F-scale items. Others may be simply agreeing to statements of this sort and the fact that some of them are "authoritarian" and others are not is a reflection of their value confusion and not necessarily indicative of potentially fascist inclinations at the personality level.

The possible confusion that might result from treating all individuals with high F-scale scores as equivalent in a study relating the F scale to a measure of intolerance of ambiguity, for example, should be obvious. Those who are high scorers by virtue of being acquiescent are by definition tolerating a considerable amount of ambiguity since they are agreeing to psychologically incompatible statements! That undue reliance on the F scale alone as a measure of authoritarianism might be misleading can be indicated by examining the pattern of responses of those undergraduates in the Northeastern sample who accepted eight or more of the ten F-scale items. Twenty of the 37 turn out on inspection to be significantly acquiescent but not significantly high on the overall measure of F, eleven are significantly authoritarian but not acquiescent, and seven are not significantly high on either measure (one individual with a score of 15 on F and a + 4 discrepancy score was both significantly authoritarian and significantly acquiescent). A virtue of the suggested method of scoring is that it gives two scores to an individual—one a measure of authoritarianism in which response set is balanced out and one a measure of response set. With a large enough number of items, it is possible to have overlap on the two

Edn
Dated
Accs No

measures, e.g., two *S*s among the 72 per cent of the graduate student sample who were significantly low on the overall measure of authoritarianism were also significantly acquiescent and another two low scorers were significantly negativistic.

It is not news, of course, to point out that the F scale is not unidimensional. The present argument suggests, however, that there are possibly greater differences among high scorers than low scorers in the customary undergraduate samples. When relatively high and low scorers are compared differences will be found when the low scoring *S*s are different from all other groups. This may account for many of the correlations found with other attitude scales when the generally more sophisticated low scorers differ from both the consistent high scorers and the confused acquiescent high scorers. The generally murky picture that prevails, however, as to the relationship between the F scale and measures of such behavioral tendencies as intolerance of ambiguity and rigidity might arise from the fact that the low scorers are similar to consistent high scorers in having an organized *Weltanschauung* but are more similar to acquiescent high scorers in being able to tolerate certain kinds of ambiguity.

These are speculations and their validity can be determined only by empirical studies using techniques that clarify the meaning of the F-scale scores of a given *S*. It must be emphasized that the majority of the samples reported on in this discussion were accidentally and not randomly drawn so the only generalization about the distribution of response set which is justified is to the nationwide population of adults with some college training where no significant amount of acquiescence was found. The presence of some acquiescence among college undergraduates raises questions about the meaning of some of the results found with other samples of undergraduates but does not necessarily imply that they too are laden with acquiescent individuals.

It should also be noted that the possibility of response set in samples studied by means of the F scale does not necessarily indicate the relevance of response set to studies where the E scale was used as the independent variable. A study on response set by Shelley yielded a correlation of $+.27$ ($p = .05$) between the anti-Semitism subscale and the use of "like much" responses to 60 pictures on the Perceptual Reaction Test of Berg and Hunt (25). This is a lower correlation than the majority of those cited earlier between the F scale and various measures of response set. Prentice found 16 favorably worded items toward Jews and Negroes correlated less highly ($+.78$) than did 16 unfavorable statements which correlated $+1.00$ with the E scale. Although we are in basic agreement with Prentice's conclusion that the use of positive or negative items is of little relevance in the measure of ethnic prejudice, we note as do Gage *et al.* that it does make some difference, a conclusion in agreement with the earlier research of Campbell and McCandless (23). Actually, this slight difference is somewhat surprising since some of the items used by Prentice violate standards we have suggested earlier, e.g., "*No* Jews are especially loud, noisy and overbearing" (23, Footnote 3, p. 420, italics added).

Our interpretation is that in these two studies, slight evidence of response set was found but, given the difficulties of comparing studies using different techniques and samples, that the amount of response set is even less than that found in studies using the F scale. This result is consistent with the general thesis that response set is greatest in ambiguous situations; the use of a minority group as a referent in an attitudinal statement is presumably less vague for most respondents than people in general or the sex life of the ancient Greeks and Romans.

In concluding, the data indicate that acquiescence is not identical with authoritarianism. This interpretation is at variance with the earlier studies reviewed, a discrepancy here attributed at least partially to the fact that many of the reversed items used in them were irrelevant to a meaningful measure of authoritarianism. It also appears that the nature of the samples used was a contributing factor since the high scoring subjects in many of these studies were inconsistent in their responses partly because they were drawn from populations containing a relatively large proportion of individuals who were, although probably but temporarily, ideologically confused.

SUMMARY

Four recent studies which have found a lack of positive correlations between the F scale

and scales of reversed items were examined. It was argued that the attempts at reversals were often characterized by items that were not psychologically opposed to the original items. It also appeared that calling upon the operation of an acquiescent response set as an explanation of the results was questionable since many of the "high" scorers in these samples were inconsistent in their responses and were not generally agreeing with F-scale items.

A new set of items designed to overcome some of our objections was developed. An item analysis indicated that the majority of the reversed items discriminated positively between individuals high and low on the F scale in two pairs of subsamples. An examination of the response patterns on the most discriminating items indicated that a minority of the respondents showed response set and that they were among the individuals who showed slight agreement with the F-scale items.

Further results were obtained on eight additional samples. All correlations were positive although there was great variability. A method for determining the significance of response set was developed and data from an additional representative nation-wide sample of adults with a college background indicated no significant degree of acquiescence. This was also true of our most sophisticated samples but appreciable amounts of acquiescence were found in an undergraduate sample. Possible implications of this finding for research using the F scale on undergraduate samples were noted. It is concluded that the data do not support the identification of acquiescence with authoritarianism.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The Authoritarian Personality*, New York: Harper, 1950.
2. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.
3. BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296-299.
4. BASS, B. M. Reply to Messick and Jackson's comments on authoritarianism or acquiescence. *J. abnorm. soc. Psychol.*, 1957, **54**, 426-427.
5. CAREY, GLORIA, ROGOW, A. A., & FARRELL, CALISTA. The relationship between the F scale and aphorism usage and agreement. *J. Psychol.*, 1957, **43**, 163-167.
6. CAMPBELL, D. T., & MCCANDLESS, B. R. Ethnocentrism, xenophobia, and personality. *Human Relat.*, 1951, **4**, 185-192.
7. CHAPMAN, L. J., & CAMPBELL, D. T. Response set in the F scale. *J. abnorm. soc. Psychol.*, 1957, **54**, 129-132.
8. CHRISTIE, R., & GARCIA, J. Subcultural variation in authoritarian personality. *J. abnorm. soc. Psychol.*, 1951, **46**, 457-469.
9. COHN, T. S. The relation of the F scale to a response set to answer positively. *Amer. Psychologist*, 1953, **8**, 335. (Abstract)
10. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475-494.
11. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3-31.
12. GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The psychological meaning of acquiescent set for authoritarianism. *J. abnorm. soc. Psychol.* 1957, **55**, 98-103.
13. JACKSON, D. N., & MESSICK, S. J. A note on 'ethnocentrism' and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132-134.
14. JACKSON, D. N., MESSICK, S. J., & SOLLEY, C. M. How 'rigid' is the 'authoritarian'? *J. abnorm. soc. Psychol.*, 1957, **54**, 137-140.
15. JONES, M. B. Authoritarianism and intolerance of fluctuation. *J. abnorm. soc. Psychol.*, 1955, **50**, 125-126.
16. JACOB, P. E. *Changing values in college.* New Haven: Hazen Foundation, 1956.
17. LEAVITT, H. J., HAX, H., & ROCHE, J. H. Authoritarianism and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215-221.
18. LORGE, I. Gen-like: halo or reality. *Psychol. Bull.*, 1937, **34**, 545-546. (Abstract)
19. MEER, S. J. Authoritarian attitudes and dreams. *J. abnorm. soc. Psychol.*, 1955, **51**, 74-78.
20. MESSICK, S. J., & JACKSON, D. N. Authoritarianism or acquiescence in Bass's data. *J. abnorm. soc. Psychol.*, 1957, **54**, 424-425.
21. NEWCOMB, T. M. *Personality and social change.* New York: Dryden, 1943.
22. PERLMUTTER, H. J. Some characteristics of the xenophilic personality. *J. Psychol.*, 1954, **38**, 291-300.
23. PRENTICE, N. M. The comparability of positive and negative items in scales of ethnic prejudice. *J. abnorm. soc. Psychol.*, 1956, **52**, 420-421.
24. PROTHRO, E. T., &. MELIKIAN, L. H. The California public opinion scales in an authoritarian culture. *Publ. Opin. Quart.*, 1953, **17**, 353-362.
25. SHELLEY, H. P. Response set and the California attitude scales. *Educ. psychol. Measmt*, 1956, **16**, 63-67.
26. WEBSTER, H. Some quantitative results. *J. soc. Issues*, 1956, **12**, (4), 29-43.

Received March 7, 1957.

DIFFERENTIAL AUTONOMIC RESPONSIVENESS IN THE PROCESS-REACTIVE CLASSIFICATION OF SCHIZOPHRENIA[1]

GERALD F. KING

Michigan State University

Considerable research attention has recently been focused on the relationship between autonomic nervous system (ANS) activity and schizophrenic behavior. Funkenstein and collaborators (9, 10, 11, 12) have been able to predict prognosis to electric shock and other forms of therapy with a variety of neuropsychiatric patients on the basis of ANS reactivity to certain drugs. Confirmatory evidence for the relationship between ANS activity and prognosis in schizophrenia has been provided in studies by Geocaris and Kooiker (14) and Hirschstein (15). In the initial method of analyzing ANS activity (9), changes in the systolic blood pressure after adrenergic stimulation (intravenous epinephrine) and cholinergic stimulation (intramuscular mecholyl) were classified into seven patterns of response. The number of blood pressure patterns has been reduced from seven to three in subsequent research (23). Further modifications have been introduced in regard to mecholyl; response to this drug has been expressed in quantitative terms, with measures of the extent and duration of blood pressure changes being employed (13, 24).

In attempting an integration of physiological and psychological processes, Meadow and Funkenstein (23) found a relationship between the ANS response patterns of schizophrenics and impairment in abstract thinking. Such a relationship was predictable on the basis of earlier studies by Bolles *et al.* (4) and others, who concluded that poor prognosis was associated with low performance on tests of abstract thinking. An investigation by Ficca (8) revealed that neuropsychiatric patients with "unfavorable" ANS patterns of response, in terms of prognosis, produced significantly more "schizophrenic" signs on the Rorschach than patients with "favorable" ANS responses. In a later study by Meadow and his associates (24), in which only mecholyl stimulation was used, measurements of ANS activity were obtained on what might be called a "level of autonomic responsiveness" dimension (e.g., maximum fall in systolic blood pressure). A significant positive relationship was found between level of autonomic responsiveness and level of abstract thinking with schizophrenic subjects. The authors interpret the results as suggesting two polar types of schizophrenia, with the descriptions corresponding closely to the so-called "process" and "reactive" syndromes. The purpose of the present research was to explore further the relationship between level of autonomic responsiveness and the process-reactive classification.

The following gives a brief delineation of the process-reactive concept, which is usually viewed as a dichotomy: process schizophrenia is indicated by a gradual onset of psychosis and a relatively inadequate prepsychotic personality, whereas reactive schizophrenia is characterized by sudden onset of illness in a relatively normal prepsychotic personality. Additional criteria are given by Kantor, Wallner, and Winder (18). The distinction, for a particular patient, is commonly made on the basis of personal history information.

A number of investigators (17, 20, 25, 27) have presented evidence showing that schizophrenics with process characteristics have lower rates of recovery than patients more reactive in nature. In a study by Kantor, Wallner, and Winder (18), the Rorschach protocols of process schizophrenics were judged as being more "psychotic" than those of reactive schizophrenics. Brackbill and Fine (6) found that process schizophrenics could not be differentiated from organics on the basis of Rorschach indices of organicity, but that reactive schizophrenics showed significantly fewer signs of organic involvement than both groups. Becker (3), conceiving of a process-reactive *continuum*, found a significant relationship between this continuum and "genetic" scores derived from the Rorschach. The data also revealed a relationship between the process-reactive dimension and the conceptual level of the responses to a proverbs test, but the relationship was restricted to male patients, not appearing in females. It can be summarized that, in comparison with process schizophrenics, reactives (*a*) have a more favorable prognosis, (*b*) produce Rorschach records which are less apt to be judged as "psychotic," (*c*) respond with fewer organic signs on the Rorschach, (*d*) give responses of a higher genetic level on the Rorschach, and (*e*) possibly interpret proverbs on a higher conceptual level.

---

[1] This research was initiated while the author was on the staff of Veterans Administration Hospital, Battle Creek, Michigan. Grateful acknowledgment is due David W. Merrell for his early collaboration, Harry VanderKamp for providing most of the physiological measures, and Frank Restle for his interest and suggestions.

The available physiological and psychological evidence suggests that the process-reactive concept can be used as a frame of reference for predicting ANS responsiveness. The following hypothesis is offered: Predominantly reactive schizophrenics exhibit a higher level of autonomic responsiveness after the injection of mecholyl than predominantly process schizophrenics. The modifying adverb "predominantly" makes the hypothesis more congruent with Becker's method of analysis (3). The variety of criteria used in making the process-reactive distinction permits patients to display a mixture of both characteristics. What is involved is a typology of personal history incidents but not a typology of patients. Thus, a given patient may fall anywhere on a scale from "purely process" to "purely reactive." To be presented in this paper are two studies which bear on the hypothesis indicated.

STUDY I

Subjects

The basic group of *S*s consisted of 136 recently hospitalized neuropsychiatric patients, all being given the diagnosis of schizophrenia. A variety of diagnostic subtypes and symptoms were represented. None of the patients showed any evidence of neurological complications from routine examination. Ages ranged from 20 to 42 years. All patients had been hospitalized eight weeks or less when the physiological measures were obtained.

Physiological Measures

The method of securing the physiological measures of ANS activity was a modification of that used by Funkenstein, Greenblatt, and Solomon (9). While the patient was lying in bed shortly after awaking in the morning, the resting systolic blood pressure was determined by a cuff manometer. The patient then received 10 mg. of mecholyl intramuscularly, and the systolic blood pressure was recorded at intervals of 2, 5, 7, 10, 12, 15, and 20 minutes. The following two measures of level of autonomic responsiveness, as derived from Meadow *et al.* (24), were obtained: (*a*) maximum fall in systolic blood pressure below the resting blood pressure following the injection of mecholyl and (*b*) area under the base-line (projection of the resting blood pressure) demarcated by blood pressure readings following mecholyl. A graphic illustration of these measures is given in Fig. 1.

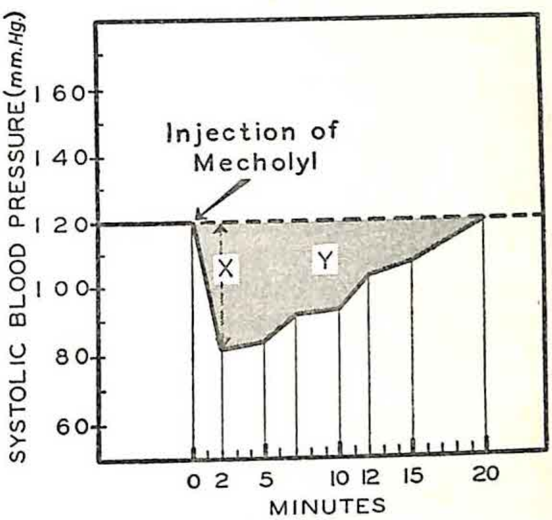


FIG. 1. DIAGRAMMATIC ILLUSTRATION OF RESPONSE OF SYSTOLIC BLOOD PRESSURE TO MECHOLYL (ADAPTED FROM MEADOW *et al.* [24])

Maximum fall in blood pressure is indicated by X. Area under baseline of resting blood pressure is indicated by Y.

The correlation between the two measures of autonomic responsiveness was found to be .82 ($N = 28$), indicating a high degree of commonality. Of the two measures, maximum fall in systolic blood pressure (MFBP) was much easier to tabulate, involving merely the subtraction of the lowest systolic blood pressure reading from the basal reading. In the present research, only MFBP was employed as a measure of level of autonomic responsiveness. Indirect evidence of the reliability of this measure is indicated by its being positively correlated with level of abstract thinking in two independent samples (24).

Procedure

A sample of 60 was drawn from the basic group of schizophrenic *S*s. All *S*s were classified on the basis of case histories as either "process" or "reactive" by the present investigator, using the frame of reference provided by Kantor, Wallner, and Winder (18). Independent classification of all *S*s was accomplished by either one of two other raters. None of the raters had knowledge of the physiological measures. The percentage of agreement between the two judgments was 76.7. Analysis of the agreement between the judgments with a fourfold table yielded a chi square of 17.47, which is significant at the .001 level of confidence. This is essentially of the same order of interrater agreement reported previously (18). Discussion among the raters revealed that at least two factors probably lowered reliability: (*a*) instances of inadequate case histories and (*b*) case histories that indicated process and reactive characteristics to an almost equal degree.

Using only *S*s on which there was classificatory agreement resulted in process and reactive groups consisting of 22 and 24 *S*s, respectively. A comparison of the groups revealed no significant difference in age. Further, in considering the process-reactive concept as a continuum, 16 *S*s were randomly selected from the two groups and ranked accordingly by two independent raters. Rank 1 was assigned to the *S* judged to be the most process in nature, with rank 16 being occupied by the *S* being most clearly reactive. Agreement between the raters is indicated by a rho of .89.

Results

The mean MFBP score of the reactive schizophrenics was 24.41 ($SD = 8.25$), as compared to 18.50 ($SD = 7.71$) for the process schizophrenics.[2] The usual t test gives a value of 2.51, which indicates that this difference is significant at the .01 level (one-tailed test). For the 16 *S*s, the rho between the sets of ranks on the process-reactive dimension and MFBP is $-.58$ ($p < .01$, one-tailed test). In accordance with the hypothesis, the reactive schizophrenics displayed a higher level of autonomic responsiveness following the injection of mecholyl than the process schizophrenics. The results also indicate that it is feasible to order the process and reactive syndromes on a continuum.

STUDY II

Subjects

With the elimination of the 46 *S*s who participated in Study I, the basic group of *S*s was reduced to 90.

Physiological Measures

The same method as in Study I was employed.

Procedure

The following modification was introduced in the classificatory procedure: all 90 *S*s were classified as either "process," "process-reactive," or "reactive." The process-reactive category was designed to be an intermediate one, representing schizophrenics with both types of characteristics to an almost equal degree. Otherwise, the same rating procedure was followed as in Study I. The percentage of agreement between the independent judgments was 87.8. The corresponding chi square, using a 3-by-3 contingency table, was 121.53, which is significant considerably beyond the .001 level. Groups representing the three categories were formed from the *S*s on which there was classificatory agreement. Balanced for *N*, each group consisted of 24 *S*s. A comparison for age yielded no significant differences among the groups.

In addition, scores on the Elgin Prognostic Scale (31) were obtained for 22 *S*s. This scale consists of ratings on factors which are based on the process-reactive differentiation. Data for the ratings were secured from interviews and clinical folders (case histories, etc.). Higher scores on the scale represent a greater incidence of process elements. Wittman (31) reports an interrater reliability coefficient of .87 for an *N* of 61. In the present research, 12 *S*s were rated independently by two raters, the interrater rho being .94.

Results

Treating the MFBP scores of the three groups by analysis of variance yielded an F ratio that is significant beyond the .01 level. The mean MFBP scores were of ascending magnitude, with the process schizophrenics being lowest ($M = 17.04$), the process-reactives intermediate ($M = 22.79$), and the reactives highest ($M = 26.62$). This trend is in accordance with the hypothesis and the conception of a process-reactive continuum. Since this is not a *post hoc* analysis, there is justification for computing t tests between the individual means. Using one-tailed tests, the results show that the process schizophrenics were significantly lower than the process-reactives ($t = 2.39$, $p < .025$), who in turn approached being significantly lower than the reactive schizophrenics ($t = 1.54$, $p < .07$). The rho between the Elgin scale and the MFBP scores ($N = 22$) was $-.49$ ($p < .025$, one-tailed test). The findings thus indicate that the reactive schizophrenics exhibited a greater autonomic response to mecholyl than the process schizophrenics; or, in other words, the process-reactive dimension is inversely related to level of autonomic responsiveness. As in Study I, the hypothesis is supported.

DISCUSSION

The results provide a basis for discussing and speculating about some of the past research dealing with physiological aspects of schizophrenia. Citing a variety of supporting studies, Angyal, Freeman, and Hoskins (2) and Hoskins (16) characterize schizophrenics, in comparison with normals, as showing reduced physiological responsiveness. However, confirmation of this position has not been obtained in a number of other studies (1, 7, 26). As suggested by Sands and Rodnick (29) and as illustrated by Malmo, Shagass, and Smith (22), contradictory results in this area are undoubtedly due in part to variations in the types of experimental stimulation and the nature of the physiological measures. Variations in the schizophrenic samples employed in different studies is another likely factor. According to the present research, responsiveness is related to the process-reactive make-up of the schizophrenic sample. If we consider pure cases of "process" and "reactive" patients for the sake of discussion, the evidence points to diminished physiological responsiveness in process but not in reactive schizophrenia. Thus, reduced responsiveness would not be found in studies where the schizophrenic subjects were predominantly reactive in character.

---

[2] In comparison with the distribution reported by Funkenstein *et al.* (13), the MFBP scores obtained in the present research seem to be depressed.

One possible reaction to the preceding discussion is that it moves in the direction of a two-factor theory of schizophrenia. Process characteristics are associated with a physiological "impairment," at least implicitly. What remains, it would seem, is to tie the disturbance in the reactive syndrome to interpersonal factors. This notion is compatible with the contention of Brackbill (5), who, after reviewing the studies of brain dysfunction in schizophrenia, postulates the existence of central nervous system pathology in process but not in reactive patients. In this context, the question arises as to whether autonomic disturbances are primary or secondary to cortical defect.

If the information on normal responsiveness is given more consideration, it is possible to cast the problem in a different light. Although the data on the reactivity of normals to mecholyl are meagre (10, 14), it seems to indicate that their blood pressure responses are very similar to what can be expected from process schizophrenics. In graphic terms, if level of autonomic responsiveness represents the ordinate and normals, reactives, and process schizophrenics lie on the abscissa in that order, the relationship can be described as curvilinear. In accordance with this hypothetical graph, the reactives would be high in autonomic responsiveness, whereas the normals and process patients would be relatively low. Such a relationship offers a more adequate account of studies which have found that schizophrenics, especially patients in the early stages of illness, are physiologically even more reactive than normals (e.g., 21, 30). A study by Pfister (28) gives a more complete illustration of the curvilinear relationship: Early schizophrenics were hyper-reactive on cardiovascular measures; if the illness persisted, however, the same *S*s dropped toward or even below the normal level. Although a loose interchange of concepts is sometimes necessary, a good deal of the available data can be fitted to the curve. A possible explanation lies in viewing both normals and process schizophrenics as representing relatively uniform, stable adjustments over a period of time. It is suggested that the corresponding physiological homeostasis is indicated by a low autonomic responsiveness. A heightened autonomic reactivity then would reflect a disruption of homeostasis, which corresponds with the more transient, precipitous nature of the reactive adjustment.

Although problems of interpretation are posed, the results in general provide additional evidence of the meaningful nature of the process-reactive classification, or dimension, of schizophrenia. Using this concept of schizophrenia, which cuts across symptomatology, evidence has been compiled from this and other studies revealing interrelationships among several classes of variables: case history information, psychological dimensions, physiological measures, and prognosis. The numerous failures of past studies dealing with schizophrenia seem to give this classification even more significance. When samples of undifferentiated schizophrenics, or even ones based on the standard diagnostic subtypes, have been employed in research, the findings have typically been either negative, not readily interpretable, or not reproducible (19). Pending the development of a better schema for delineating the heterogeneous category of schizophrenia, the process-reactive classification seems to offer a productive orientation for research.

SUMMARY

This investigation consisted of two studies which were concerned with the relationship between the process-reactive classification of schizophrenia and autonomic nervous system activity. On the basis of psychological and physiological findings, it was hypothesized that predominately reactive schizophrenics display a higher level of autonomic responsiveness after the injection of mecholyl than predominantly process schizophrenics. Similar in nature, the two studies had the following methodological features in common: (*a*) The process-reactive classification of schizophrenic *S*s was made on the basis of case history information, (*b*) a small sample of the total *N* was ranked on a process-reactive continuum, and (*c*) maximum fall in systolic blood pressure constituted the measure of autonomic responsiveness.

In accordance with the hypothesis, the results of both studies revealed that the patients classified as "reactive" exhibited a significantly greater fall in blood pressure after the administration of mecholyl than the process patients. Significant rank-order correlations were also obtained between the process-reactive dimensions employed in each study and maximum fall in blood pressure, indicating that it is feasible to order the process and reactive syndromes on a continuum. Some implications of the findings in regard to physiological responsiveness in schizophrenia were discussed. It was concluded that the process-reactive classification seems to provide a fruitful frame of reference for research in schizophrenia.

REFERENCES

1. Altschule, M. D., & Sulzbach, W. M. Effect of carbon dioxide on acrocyanosis in schizophrenia. *Arch. Neurol. Psychiat.*, 1949, **61**, 44–55.
2. Angyal, A., Freeman, H., & Hoskins, R. G. Physiologic aspects of schizophrenic withdrawal. *Arch. Neurol. Psychiat.*, 1940, **44**, 621–626.
3. Becker, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229-236.
4. Bolles, M. M., Rosen, G. P., & Landis, C. Psychological performance tests as prognostic agents for the efficacy of insulin therapy in schizophrenia. *Psychiat. Quart.*, 1938, **12**, 733-737.
5. Brackbill, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, **53**, 210–226.
6. Brackbill, G. A., & Fine, H. J. Schizophrenia and central nervous system pathology. *J. abnorm. soc. Psychol.*, 1956, **52**, 310–313.
7. Cameron, D. E. Heat production and heat control in the schizophrenic reaction. *Arch. Neurol. Psychiat.*, 1934, **32**, 704–711.
8. Ficca, S. C. Relationship of "autonomic" blood pressure pattern types and subjects' performance on the Wechsler-Bellevue and the Rorschach test. Unpublished doctoral dissertation, Pennsylvania State Univer., 1950.
9. Funkenstein, D. H., Greenblatt, M., & Solomon, H. C. Autonomic nervous system changes following electric shock treatment. *J. nerv. ment. Dis.*, 1948, **108**, 409-422.
10. Funkenstein, D. H., Greenblatt, M., & Solomon, H. C. Psycho-physiological study of mentally ill patients. *Amer. J. Psychiat.*, 1949, **106**, 16-28.
11. Funkenstein, D. H., Greenblatt, M., & Solomon, H. C. Autonomic changes paralleling psychologic changes in mentally ill patients. *J. nerv. ment. Dis.*, 1951, **114**, 1–18.
12. Funkenstein, D. H., Greenblatt, M., & Solomon, H. C. An autonomic nervous system test of prognostic significance in relation to electroshock treatment. *Psychosom. Med.*, 1952, **14**, 347-362.
13. Funkenstein, D. H., Greenblatt, M., & Solomon, H. C. Nor-epinephrine-like and epinephrine-like substances in psychotic and psychoneurotic patients. *Amer. J. Psychiat.*, 1952, **108**, 652–662.
14. Geocaris, K. H., & Kooiker, J. E. Blood pressure responses of chronic schizophrenic patients to epinephrine and mecholyl. *Amer. J. Psychiat.*, 1956, **112**, 808–813.
15. Hirschstein, R. The significance of characteristic autonomic nervous system responses in the adjustment, change and outcome in schizophrenia. *J. nerv. ment. Dis.*, 1955, **122**, 254–262.
16. Hoskins, R. G. *The biology of schizophrenia.* New York: Norton, 1946.
17. Kant, O. Differential diagnosis of schizophrenia in light of concepts of personality-stratification. *Amer. J. Psychiat.*, 1940, **97**, 342-357.
18. Kantor, R. E., Wallner, J. M., & Winder, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157-162.
19. King, G. F. Research with neuropsychiatric samples. *J. Psychol.*, 1954, **38**, 383–387.
20. Langfeldt, G. The diagnosis of schizophrenia. *Amer. J. Psychiat.*, 1951, **108**, 123–125.
21. Malmo, R. B., & Shagass, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11**, 9-24.
22. Malmo, R. B., Shagass, C., & Smith, A. A. Responsiveness in chronic schizophrenia. *J. Pers.* 1951, **19**, 359–375.
23. Meadow, A., & Funkenstein, D. H. The relationship of abstract thinking to the autonomic nervous system in schizophrenia. In P. Hoch & J. Zubin (Eds.), *Relation of psychological tests to psychiatry.* New York: Grune & Stratton, 1952. Pp. 131-149.
24. Meadow, A., Greenblatt, M., Funkenstein, D. H., & Solomon, H. C. Relationship between capacity for abstraction in schizophrenia and physiologic response to autonomic drugs. *J. nerv. ment. Dis.*, 1953, **118**, 332–338.
25. Milici, P. Postemotive schizophrenia. *Psychiat. Quart.* 1939, **13**, 278–293.
26. Parsons, E. H., Gildea, E. F., Ronzoni, E., & Hulbert, S. Z. Comparative lymphocytic and biochemical responses of patients with schizophrenia and affective disorders to electroshock, insulin shock, and epinephrine. *Amer. J. Psychiat.*, 1949, **105**, 573–580.
27. Paskind, H. A., & Brown, M. Psychoses resembling schizophrenia occurring with emotional stress and ending in recovery. *Amer. J. Psychiat.*, 1940, **96**, 1379-1388.
28. Pfister, H. O. Disturbances of the autonomic nervous system in schizophrenia and their relations to the insulin, cardiazol and sleep treatments. *Amer. J. Psychiat. Suppl.*, 1938, **109**, 94–118.
29. Sands, S. L., & Rodnick, E. H. Concept and experimental design in the study of stress and personality. *Amer. J. Psychiat.*, 1950, **106**, 673–679.
30. Williams, M. Psychophysiological responsiveness to psychological stress in early chronic schizophrenic reactions. *Psychosom. Med.*, 1953, **15**, 456–462.
31. Wittman, Phyllis. Scale for measuring prognosis in schizophrenic patients. *Elgin State Hosp. Papers*, 1941, **4**, 20–33.

Received May 31, 1956.

AMBIGUITY AND ANXIETY[1]

ANDREW S. DIBNER

Clark University

THE concept of anxiety has been subject to considerable scrutiny in the last few years. The present study is concerned with the meaning of anxiety, where anxiety is considered as a momentary state, not as a "personal constant" (14) or as a disease process (17). The thesis presented is that anxiety is directly related to the degree of ambiguity in the situation to which the individual must make some adjustive reaction, that it is the condition of *uncertainty* (of perception, expectation, success, prediction, control, or even survival) which is the essence of anxiety, and that the probability of uncertainty is greater whenever external conditions are ambiguous. The ambiguous field permits a variety of possible structures (20), offering the individual a complex array of choices with a lessened probability of correctly predicting the results of his decisions because of a lack of definitive cues. Cues may be absent, vague, and inconsistent or contradictory. A highly structured field, on the other hand, provides guide lines; the individual may properly evaluate it and behave toward it in a meaningful way. There is no implication that ambiguity is always a sufficient condition for anxiety; the present argument bears only on situations containing some element of personal importance or ego involvement and calling for an adjustive response.

In this paper the terms "ambiguity" and "structure" are used to describe both objective properties of the stimulus field and psychological interpretations of the field. The stimulus field is said to have greater structure according to the extent to which it contains sufficient perceivable characteristics to be subject to a greater consensus of interpretations by observers. It is relatively ambiguous (or unstructured) if its characteristics are less sharply defined, permitting more alternative interpretations. Thus, one can diminish the structure of a stimulus field in an objective way by subtracting information from it, or one can add to structure (diminish ambiguity) by adding information. On the other hand, the psychological interpretation of the structure of a stimulus field (Subjective Structure as opposed to Objective Structure) has as its determinants not only the structural properties of the field but also those intrapersonal attributes which affect perception. Persons differ in their tendency to impose structure on an ambiguous field (1, 15, 24), and it seems probable that even a rather highly structured stimulus may be perceived as ambiguous by some.

There have been few experiments related to such hypotheses about the relationship of anxiety to objective and subjective structure (7, 13, 28) and none in a natural field situation. This study utilized five measures of anxiety to test patients' reactions to psychological interviews which were varied in the degree of structuring provided by the interviewers.

METHOD

Subjects

The 40 *S*s were male veterans selected from consecutive admissions to the Neuropsychiatric Service of the Dearborn Veterans Administration Hospital.[2] All cases of suspected psychosis or organic brain disease were excluded, as were those involving poor verbal ability or low IQ. Selected *S*s were all cases of neurosis, psychosomatic disease, or psychopathic personality. They had to be ambulatory and in sufficient control to be on the open wards, relatively new arrivals in the hospital, no more than 55 years of age, and without previous contacts with the Clinical Psychology section before the experimental interview. All were seen within one week of admission and were randomly assigned to interviewer and conditions. A later check revealed equivalence of the groups of *S*s receiving Structured or Ambiguous interviews on such variables as age (Range 25–54, Mean 31.7), diagnosis, race, education (Mean 10.5 years), IQ (Mean 105.4), marital status, number of days in the hospital before being interviewed, sedation on the day of the interview, time of day of interview, and the number and type of previous psychiatric or psychological contacts.

---

[1] This paper is based upon a doctoral dissertation done at the University of Michigan (9). The writer wishes to express his appreciation to Edward S. Bordin for his stimulating guidance. Some support was received from the Research Project: Analysis of Therapeutic Interaction, USPHS M-516, National Institute of Mental Health, Bethesda, Maryland.

[2] The writer is indebted to Thomas P. Crane, Manager, and J. J. Lasky, then Chief Clinical Psychologist, for their help.

Procedure

Four psychologists[3] each conducted clinical evaluative interviews, five in an ambiguous manner and five in a structured manner, thus creating a group of 20 Ambiguous (A) and 20 Structured (S) interviews. The interviews were tape recorded and readings were made of the patients' skin resistance throughout the interviews. Following the interviews, the patients were asked standardized questions by *E*, and completed check lists and cardsorts to obtain their reactions to the sessions. Five anxiety measures were compared with an objective index of structuredness of interviewer behavior and with a measure of patients' perception of structure.

Conditions. Situational factors of the interview setting were strictly controlled. *S*s were not told that the interviews were experimental; the recording and skin resistance apparatus were explained as standard hospital techniques. The post-interview data collecting, however, was justified to the *S*s as a research study. The interviewers were trained in the two conditions. In the A condition, the patient was told it was considered best if he talked about himself in any way he wished, saying whatever he thought would be of help to the interviewer in understanding him and his life. The interviewer accepted anything brought up and supported when necessary, but tried to avoid giving any but the most general cues as to what to talk about. He tried to avoid giving advice or asking follow-up questions. In the S condition, the interviewer asked questions in predetermined areas and followed up any lead the patient offered. He was free to make his observations or feelings about the material known to the patient in the form of questions, information, or counsel whenever he wished.

Measures of Structure

Objective Structure. The degree of structuring by the interviewers was measured by applying Osburn's method (23) to 10-minute samples of the typescripts. This index of degree of structure in an interview is based on a weighted coding of individual interviewer responses and on the total number of his responses. The coding considers interviewer responses in terms of their direction (information-giving or information-seeking), referent of the response (client referent or situation referent) and generality of the response (most general to highly specific in content). Since the distribution of these scores was considerably skewed, it was necessary to use nonparametric statistics in the analysis of results.

Subjective Structure. Subjective structure was defined by clinical judges' ratings of the degree of ambiguity as seen by the patient. Ratings were based on analysis of transcripts of patients' responses to three open-ended questions following their interviews. The questions asked how the patient felt the interview went, how it could have gone better, and if he felt he knew what to talk about in the interview. Instructions to raters follow:

> Following a clinical first interview with a psychologist, patients were asked about it by the researcher. Certain questions asked them were designed to elicit *their perception* of the degree of structuredness or ambiguity of the interview situation they had just faced. Degree of structuredness of the interview *as experienced by the patient* is what is being considered now.
>
> The patients had been faced with the task of telling the psychologist about themselves. What they should or should not tell the psychologist is determined by many factors. For example, the degree of defining or structuring by the interviewer is a factor. Another factor involves the ideas the patient has about what is important in his case and what he thinks the interviewer is interested in. Also the *needs* the patient might have to reveal some things and keep others secret would affect his perception of what should be talked about. Thus, with some factors coming from within the patient, his *perception* of the structure of the task might be different from the perception other patients might have facing the same interviewer in the same situation. Even though an interviewer may feel he is providing many cues as to what the patient should talk about, some patients might still feel as if the situation is very ambiguous. Others might see their interview as cut-and-dried, highly structured, even though their interviewer doesn't ask them one specific question. They might have a structure of their own that they apply.
>
> The responses to certain questions by nine patients are presented here. Please rate for each patient how you think he perceived the *structuredness* of the interview he had. Rate each interview by placing check marks on the 6-point linear scales on the rating sheet. You can check between numbers if unsure of the placement.

---

[3] The author is grateful to Richard Cutler, Barbara Nachmann, Stanley Segal, and David Shapiro for their cooperation as interviewers.

____1____2____3____4____5____6____

Ambiguous. The patient felt he didn't know what to talk about in the interview he had.	Structured. He had no doubts at all concerning what he should talk about.

Rank intercorrelation of five judges rating nine patients was .56. *E* rated all 40 patients. His rank correlation with the median of the five raters was .90.

Measures of Anxiety

Palmar Skin Conductance. Resistance readings were recorded at 30-second intervals by the *E* in a room adjoining the interview room. Following an accommodation period of four minutes, a curve of conductance was plotted for each patient over a 35-minute period. The slope of the curve from the mean of the first 2½ minutes to the mean of the final 10 minutes was used as the skin conductance measure of anxiety for the interview. The mean and range of conductance was equivalent to data reported by others (28, 31). Range of conductance-change through the interviews was from −180 to 952 micromhos, with mean change of 148.5.

Anxiety Cards. Following the interview, *E* presented each *S* with 20 cards, one at a time, each containing phrases describing feelings of tension or absence of tension or common physiological concomitants of tension. The *S* was asked to place each card on one of three piles indicating that he "Felt that way," "Was not sure," or "Did not feel that way." Statements on the Anxiety Cards appear below. The starred items are nonanxious in character.

1. I felt I couldn't concentrate too well.
2. I felt I was sweating quite a bit.
3. During most of the talk I felt perfectly at ease.*
4. I felt restless most of the time.
5. I felt my stomach was "turning over".
6. I kept wondering what this interview was all about.
7. I felt clear-headed all the way through the talk.*
8. After about five minutes I felt relaxed and felt relaxed till the talk was over.*
9. I felt nervous through the whole talk.
10. The talk left me feeling tired.
11. I felt too hot.
12. I was worried about what I was telling him.
13. I felt pains in my chest sometimes while we were talking.
14. I had difficulty in breathing sometimes while we were talking.
15. I felt very comfortable from the beginning of the talk right through to the end.*
16. Sometimes during the talk I could hear my heart pounding.
17. Sometimes my throat felt dry while we were talking.
18. I felt my muscles tighten up sometimes during the talk.
19. I felt much better during this talk than when I talk to most doctors.*
20. Most of the time I felt rather tense.

These statements were selected on the basis of a pilot study using a longer list of anxiety statements. The five Nonanxious statements were included with the 15 Anxious statements to permit *S* to profess control of anxiety, but only the number of cards bearing anxious statements that the patient sorted as "Felt that way" yielded his Anxiety Card score. The split-half reliability of the scale was .76. An item analysis of the anxious cards further showed that the scale had high internal consistency. Obtained scores ranged from 0 to 11 with a mean of 3.15.

Anxiety Ratings. Ratings of anxiety were made by clinical judges from transcripts of 10-minute samples of the interviews. Four VA staff psychologists were given 17 10-minute samples of transcribed interviews and were asked to rate each patient response or 10-line segment of long responses on a 5-point scale of anxiety. The sample of 17 cases included two for each interviewer for each condition, plus one randomly selected case. Instructions to raters appear below:

> Read through the transcript and rate each patient response (or 10-line segment of long responses as indicated). Use the following 5-point scale of Anxiety:
>
> ____1____2____3____4____5____
>
> | | | |
> |---|---|---|
> | Relaxed, not anxious | Anxious | Very anxious |
>
> Try to consider the scale as linear and continuous with 3 as the center-point representing the average anxiety level common to patients having their first clinical interview. You may check between the numbers if unsure of placement of some ratings.
>
> In reading each patient statement try to get a feel for his emotional state at the time he is saying those words. Content may or may not be important. Try to judge mainly by the way he verbalizes the content. Here are some suggested cues. Also use that "3rd ear" freely to pick up the patient's emotional tone.
>
> Some signs of anxiety:
>
> Fumbling around with how to present thoughts.
> Sighing, tense laughter, changes in pitch or volume.
> Tripping over words, stuttering, leaving words or thoughts half finished.
> Patient stating he is uncomfortable or apprehensive.
> Angry tone that sounds like reacting to anxiety.
>
> Some signs of little anxiety:

Smooth running speech, thoughts connected and uninterrupted.

Statements of confidence that don't smell of defensiveness.

Some responses are very short but try to rate them as well as you can. For example: The patient may ask the interviewer to repeat (P: What?); from the context try to get a feel for what is behind this "What?" At that point the patient may be anxious or relaxed.

For each rater, the individual ratings of responses within an interview were averaged to obtain a single score for an interview. Rank intercorrelation of the four raters was .59. One of the raters, whose correlation with the median of the other three raters was .52, judged the rest of the interview samples.

Because of the possibility of bias from the rater's knowledge of the hypotheses and from his judgment as to interviewer structuring (by reading interviewer responses), a test of presence of bias was made. Three additional raters were given 16 of the interview samples with all interviewer responses, short patient responses, and any patient references to lack of structure eliminated. The three raters of these specially prepared samples had a rank intercorrelation of .60, achieving about the same degree of agreement as full sample raters. Their median ratings of the 16 interviews correlated .64 with the median ratings of the full sample raters. If corrected for attenuation due to inaccuracy of rating, this correlation would actually approach unity. The variance unaccounted for by the correlation of .64 is almost entirely a function of the unreliability of ratings, leaving little room for effect by bias on the part of the full sample raters.

Speech Cue Scales I and II. Two measures of clinical signs of anxiety in speech, scored from the 10-minute samples of the transcripts, were employed. Their method of development is reported elsewhere (8). The scales are based upon a count of speech cues reflecting disorganization, emotional involvement, and defensiveness. The cues included in each scale are as follows:

Cue Scale I

BR —Breaking in with a new thought, generally by breaking into another sentence.

UN —Unfinished sentence.

IDK—Saying "I don't know," not in answer to a direct question, but as in resignation, disgust, or despair.

INT—Interrupted sentence. The thought is interrupted within the sentence but is eventually completed in much the same manner as it was begun.

R —Repeating words or phrases when full words or phrases are repeated close in sequence within a sentence.

ST —Stuttering when repeating parts of words or when leaving a word unfinished.

Cue Scale II

L —Laughing. Any kind of laugh or chuckle.

V —Voice change. Change in volume, becoming louder or lower.

SI —Sighing or deep breath.

By inspection of a table of their intercorrelations, the speech cues were grouped into two patterns defining the two scales. With the exception of IDK, the cues of Scale I are all indicators of speech disruption. Scale II is composed of cues related to nonverbal voice characteristics. The agreement of two judges rating the cues grouped in Scale I was 65 per cent, the cues grouped in Scale II, 100 per cent. Cue Scales I and II have a Kendall's rank correlation (tau) of $-.03$ ($p = .81$). Unrelated, they were used independently as anxiety measures.

RESULTS

Structuredness

Applying the Objective Structure measure to the interviews revealed that the interviewers were able to adhere to the conditions. Out of the total group of 39 interviews of which transcriptions were made, all but two of the S interviews had higher Objective Structure scores than did the A interviews, and only two A interviews overlapped the S distribution. Interviewers were found by analysis of variance by ranks (16) not to differ significantly in adhering to the conditions. Thus, interviews by all interviewers could justifiably be pooled for analysis.

Subjective Structure or the patients' perception of structuredness was found to be strongly related to the experimental conditions ($r_{\text{bis}} = .71$), but the relationship was considerably lessened when individual variation was more fully represented by using a rank correlation with Objective Structure (tau $= .35$). Thus, a patient's perception of structure is in part determined by the structuring behavior of the interviewer, but variance appears to be introduced by the unreliability of the measures and interpatient personality differences which affect perception.

Measures of Anxiety

Intercorrelations among the five anxiety measures used as independent criteria appear

TABLE 1

CORRELATION (TAU) OF ANXIETY MEASURES (40 interviews)

	Anxiety Cards	Skin Conductance	Anxiety Ratings	Cue Scale I
Skin conductance	.14	—	—	—
Anxiety ratings	.00	−.02	—	—
Cue Scale I	.01	−.08	.32**	—
Cue Scale II	.07	−.14	.26**	−.03

** $p < .01$.

TABLE 2

CORRELATION (TAU) OF STRUCTURE MEASURES AND ANXIETY MEASURES ACROSS ALL INTERVIEWS

	Anxiety Ratings	Anxiety Cards	Skin Conductance	Cue Scale I	Cue Scale II
Objective structure	−.40**	−.19*	−.05	−.18*	−.23*
Subjective structure	−.03	−.27**	−.22*	−.03	−.04

Note.—Since direction of relationship was predicted, one-tailed test was used for significance of correlations.

* $p < .05$.

** $p < .01$.

in Table 1. Only the correlations between the Cue scales and Anxiety ratings reach significance, and these are low. Both Cue scales and the Anxiety ratings were based on the same data—transcribed patient responses. Judges doing the anxiety ratings were instructed to be aware of speech clues in their estimate of patient anxiety. Although the two Cue scales are themselves uncorrelated, both sets of speech cues enter into a judge's estimate of anxiety when reading transcripts.

Ambiguity and Anxiety

A test of the hypotheses of relationship between anxiety and ambiguity was made by correlating the measures of Objective and Subjective structure with the five anxiety measures across all interviews. Table 2 presents these correlations. In anxiety *vs.* structure of the situation as defined by the interviewer (Objective Structure) all correlations are negative, and four of the five reach significance, though all are low. The largest negative relationship (−.40) is found between structuredness of interviewer behavior and clinical judgment of patient anxiety. No significant relationship was found between Objective Structure and the autonomic measure of anxiety (skin conductance).

Similarly, correlations of anxiety to patients' perception of structure (Subjective Structure) are again all negative, but only two anxiety measures, the patient's self-report (Anxiety Cards) and his change in skin conductance, show significant relationships.

Thus, the hypotheses of negative relationship between anxiety and structure are borne out by four anxiety measures in the case of Objective Structure and two in the case of Subjective Structure.

In these analyses, the total group of experimental interviews was used because the measures of Objective and Subjective Structure could be applied to all interviews, ranging them on continua. However, the total group was, from another point of view, made up of two groups toward which the interviewers were instructed to behave differently, as to both their early orientation of the patient and their structuring behavior throughout the interview. It was thought that a more rigorous test of the hypotheses could be attempted by examining their utility *within* the separate A and S groups. Such an analysis would eliminate effects due to the difference in initial orientation of patients and the difference in expectations of the interviewers. Presumably, within each group (A or S) each interviewer tried to be *the same* as he was in other interviews in that group, not *different*, as he had to be when he interviewed the next patient under the other condition. Thus, a test of the hypotheses within probably more homogeneous groups would be a purer and stricter test. Because N was reduced to half, providing a small sample of only 20 cases (in some cases 19), and because the range of structuredness is reduced, it would be expected that error factors would have a larger effect in such an analysis, and higher correlations would be needed to reach significance. By comparing results from the analysis of these subgroups with findings across all interviews, however, one would be perhaps better able to understand the nature of variation in the interview situation.

Table 3 presents correlations of Objective and Subjective Structure with anxiety within the experimental A and S groups. All the relationships are negative within the groups, except those between Objective Structure and Skin Conductance and between Subjective

TABLE 3

CORRELATION (TAU) OF STRUCTURE MEASURES AND ANXIETY MEASURES WITHIN EXPERIMENTAL A AND S GROUPS

	Anxiety Ratings	Anxiety Cards	Skin Conductance	Cue Scale I	Cue Scale II
A Group ranked on objective structure	−.32*	−.10	.00	−.42**	−.24
S Group ranked on objective structure	−.46**	−.20	−.12	−.50**	−.12
A Group ranked on subjective structure	−.12	−.24	−.28*	−.10	.01
S Group ranked on subjective structure	.13	−.33*	−.24	−.11	−.02

Note.—One-tailed test of significance was used unless correlation was positive.

* $p < .05$.

** $p < .01$.

Structure and Anxiety Ratings and Cue Scale II, all three of which are not significantly different from zero.

Several points stand out in this table. One is that to a great extent the relationships which attained significance on the total sample (see Table 2) either attain or approach significance within the separate experimental groups. Second, the relationships do not seem equivalent in the two experimental groups examined separately. In four out of the five correlatious with Objective Structure, the S group showed somewhat larger absolute negative relationships than the A group. Third, Cue Scale I showed comparatively large correlations (−.42 and −.50) with Objective Structure within both subgroups, but when the total sample was used, the correlation shrank to −.18.

The first point, concerned with correlations between structure and anxiety within groups, is consistent with the hypotheses. The within-groups analyses apparently eliminated the artificial experimental condition of instructing interviewers to behave in a structured or ambiguous way. With only the natural variation in structuring behavior the interviewers planlessly provided, the results are still of a kind suggested by the hypotheses.

As for the second point, one interpretation is that interviews in the two experimental groups necessarily had different psychological atmospheres which were related to the experimental conditions which distinguished them, and that variation in structuring behavior by the interviewer subsequently had a somewhat different effect, depending upon this more general variable. It must be remembered that there were only two real differences between the A and S groups. One was that the interviewers were instructed to be unstructured in the A group and structured in the S; the other was that they told patients early in the A interviews that they would not structure it for them, repeating this orientation several times during the interview if necessary. Since one group (A) was given to understand there would be little structuring and the other (S) had no such orientation, one may hypothesize that subsequent variability in structuring behavior by the interviewer was differently interpreted and reacted to by patients in the two groups. The finding of somewhat larger negative correlations between anxiety and structure within the S than the A group is consistent with this *post hoc* hypothesis. Seemingly, patients whose interviewers said they were going to be unstructured and instead were fairly structured showed less anxiety than did patients whose interviewers started off with a structured approach and wound up being fairly unstructured. This is made clearer by considering the third point brought out in the table, that of rather higher correlations between Cue Scale I and Objective Structure were within groups than when the groups were combined.

Scatterplots of Cue Scale I and Objective Structure within A and S groups are presented in Fig. 1 side by side. The two cases of A whose Objective Structure overlapped the range of S interviews are circled. Similarly, the two S interviews whose Objective Structure was so low as to overlap the range of A interviews are circled. The lines connect mean points of five intervals in the distributions and were inserted to illustrate the regressions. One can clearly see the anxiety effects in the S interviews in which the interviewers gave comparatively little structuring, and the relatively low anxiety effect in the A interviews which were moderately structured. When all interviews were pooled for analysis, the correlation of Objective Structure and speech disruption

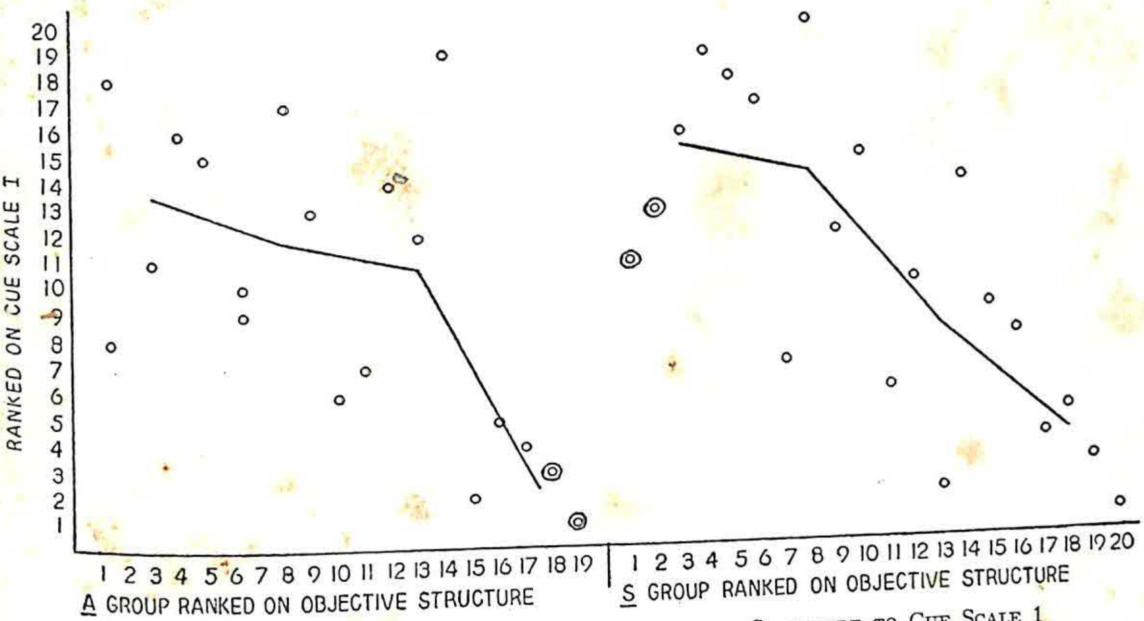


FIG. 1. SCATTERPLOT SHOWING RELATIONSHIP OF OBJECTIVE STRUCTURE TO CUE SCALE 1 WITHIN A AND S GROUPS

was lowered by these effects. This interpretation has a kind of common sense basis in the notion that any orientation offered to an *S* adds to the structure of the stimulus situation he faces, even the A orientation in which patients were told at the start of the interview that the interviewer would do little structuring!

The possibility of interviewer differences in arousing anxiety in the patients was explored by using the H test (16) over all interviews. None of these tests showed significance. Thus, there is no evidence that patients showed more anxiety with one interviewer than another. This factor, therefore, seems unimportant in evaluating present results.

Similarly, because disproportions in topical content during the interviews might influence findings on anxiety, an analysis was made of topics discussed during the 10-minute samples of all interviews. Interviews of A and S types were tallied on a classification of topical areas suggested by the content and subjected to chi square test. No significant differences emerged.

DISCUSSION

A methodological factor may have had some effect on the inconsistent relationships of the anxiety measures both with each other and with the two measures of structure. Some of the measures were "global" ones, derived either from data collected during the whole interview or from impressions after the interview was completed; other measures were derived only from data obtained during the 10-minute samples, which began 20 minutes after the start of the interview. Higher relationships were found, in general, among measures of the global type or among measures of the time-sample type than when the two types were compared. In general, the global measures of anxiety (Anxiety Cards and Skin Conductance) related better to the global index of structure (Subjective Structure) than they did to the time-sample measure. Similarly, the anxiety measures which were derived from the time sample (Anxiety Ratings and Cue Scales I and II) related better to the measure of structure derived from the same sample than they did to the global measure of structure.

There are, however, some significant relationships between time-sample measures and global measures. More importantly, Objective Structure, based on consideration of the interviewer's behavior in the 10-minute samples, was found to be highly correlated with an estimate of his behavior based on the first five minutes of the interview (rho = .85), and may thus be taken as an adequate estimate of the structuredness of the whole interview. Therefore, any relationships of other measures to Objective Structure may be interpreted as meaningful in terms of the whole interview, not just the 10-minute sample.

However, these methodological considerations do not adequately explain the lack of correlation among the anxiety measures. The two speech measures, both derived from the time sample, are uncorrelated. Similarly, Skin Conductance and Anxiety Cards, both global measures, have no significant relationship. An understanding of the lack of correlation of measures, each of which has theoretical rationale as being an index of that emotional disturbance called anxiety, sheds light on the nature of the concept of anxiety. The Anxiety Cards assess what the patient can or is willing to admit about the mental or somatic discomfort he experienced. They reflect tension of which the patient is *consciously aware*, although, of course, the degree of tension of which he is aware may be quantitatively more or less than what he indicates in his choice of Anxiety Cards. Skin Conductance, on the other hand, is one measure over which the patient has little conscious control (22). Previous studies have claimed both that *S*s are aware and unaware of changes in skin conductance (19, 32). Studies generally agree that it represents "built up" (unexpressed) tension. Perhaps there is a threshold of awareness for skin conductance changes which varies with the individual. The two uncorrelated Cue scales represent speech variables indicative of "emotionally involved" or disorganized verbal behavior. Upon analysis, some of these can be understood as direct instances of disorganization (incomplete sentence or breaking in of new thoughts) and others as defenses against or attempts to control disorganization (repetition or modification of words and phrases). Very likely, these cues are not subject to the person's awareness. The Anxiety Ratings, which are made by a clinical judge from the patient's verbal productions, can also be considered as based on a variety of cues of emotional disturbance or defense against disorganization of which the patient is essentially unaware.

The suggestion of relationship between Anxiety Cards and Skin Conductance (tau = .14) may be a reflection of the "aware" element of Skin Conductance with the "Aware-and-willing-to-admit-it" element in the Anxiety Cards measure. In any case, these low correlations force the conclusion that if all of these measures do indeed refer to anxiety, anxiety is a combination of "emotional" phenomena, each bearing little relationship to the other, some being available to the person's awareness while others, perhaps, are not although they are identifiable by an observer. Lacey and Van Lehn (18) and others (6, 21, 22) offer a theory of emotional reaction in which each individual is thought to have his own hierarchy of patterns of physiological reaction, in opposition to the view (3, 4) that certain specific emotional states evoke specific reactions in all people. Also, there is support from personality studies of psychic defense reactions that persons not only have individual hierarchies of defenses but that some use their defense system over all conflict situations, whereas others employ various reaction mechanisms, depending on the nature of the conflict situation and its relevance to the individual's unique personality dynamics (11, 12, 27, 29). This approach to the study of personality assumes that there are certain reaction mechanisms which are widely available to organisms as a group but which have varying salience, perhaps over wide temporal or situational spheres or changing with different situations, for different individuals. Applying this type of consideration to the indices of anxiety, one can imagine that each may indeed be a measure of anxiety but that individuals differ markedly in their expression of tension in speech, in autonomic changes, and in awareness of mental and somatic tension. Moreover, within the individual, there may be changes in the pattern of reaction mechanisms, depending on his threshold of tolerance for tension. Anxiety, then, may be an emotional state not subject to stable measurement by any one behavioral or physiological yardstick in any group of individuals, or in any one individual in different circumstances of stress or motivation.

Thus far, perception of structure has been considered as subject to influence by the objective nature of the situation and by the patient's anxiety. Actually, the relationship between perceived structure and patient anxiety may be a reciprocal one, the patient's report of structure being influenced to some extent by the degree to which he felt threatened by the interview. The findings in regard to Subjective Structure and anxiety are more easily understood in this light. The two anxiety measures which showed correlation with Subjective Structure were Anxiety Cards and Skin Conductance. These measures are probably most representative of the patient's awareness of tension. Frenkel-Brunswik (10) found, in this connection that those who are tolerant of ambiguity (who perceive the situation as ambiguous if it objectively is) are able to tolerate the anxiety involved, while those rigid persons who perceptually engraft structure on an ambiguous situation do so to avoid anxiety. The present findings support Frenkel-Brunswik's in that perception of ambiguity was indeed related to "feeling anxious." The findings in relation to Objective Structure, however, demonstrate that people may be anxious at a level below awareness when facing an ambiguous situation. Was such the case with Frenkel-Brunswik's intolerant group?

Finally, although the goals of therapy most often involve reduction of anxiety—not elimination, but reduction to a manageable and noncrippling level (26, 30)—theorists differ as to the value of anxiety as a therapeutic agent. Analytic theorists depend on anxiety as a motivator for progress (2, 25), whereas client-centered therapists feel that anxiety reduction must be a constant goal ". . . because only in a non-threatening situation the individual's capacity for growth is freed" (5, p. 360). In any case, this study demonstrates that interviewers can deliberately vary the degree of structure in the situation. In view of the findings of a relationship between Objective Structure and various types of anxiety, this suggests that structuring behavior can be planfully used by a clinician as a technique for the therapeutic control of anxiety when he thinks his patient has surpassed or failed to achieve some optimal anxiety level. The research implications of this possibility are clear and important.

SUMMARY

Considering anxiety in its meaning as a momentary state, not as a personality variable, this study tested the hypotheses that anxiety is related to the ambiguity of the presenting situation as objectively measured and as perceived by the subject.

Forty neuropsychiatric patients in the Dearborn VA Hospital were given clinical evaluative interviews by four interviewers. In half the interviews, the interviewer gave few clues to guide the patient in discussing his life and problems (Ambiguous condition). In the other interviews, the interviewer took an active part in guiding the discussion (Structured condition). All interviews were tape-recorded, and a record was obtained of patients' skin conductance throughout the interview. Patients were then questioned as to their reactions to the talks.

Two measures of structuredness were obtained: a measure of the structuring qualities of the interviewer's verbal behavior as analyzed from the typescripts, and a rating of the patient's perception of the structuredness of the interview situation as analyzed from his post-interview report. Five different measures of anxiety were used: changes in palmar skin conductance, ratings of anxiety by a clinical judge reading samples of the interview transcript, the patient's own report of tension by means of a modified check-list, and two different sets of indices of disturbed speech.

Four out of the five anxiety measures showed significant relationships with objectively measured structure, and two out of the five showed significant relationships with structure as the patient perceived it, partially confirming the hypotheses.

The general lack of correlation of the anxiety measures with each other and their patterns of relationship to the two measures of structure were discussed in terms of the units of information on which the measures were based, and in terms of the multidimensional nature of anxiety response mechanisms and the interrelationships of anxiety, perception, and personality structure. Findings of the study were discussed in relation to a theory of anxiety and to psychotherapy.

REFERENCES

1. ADORNO, T. W. *et al. The authoritarian personality.* New York: Harper, 1950.
2. ALEXANDER, F., & FRENCH, T. M. *Psychoanalytic therapy.* New York: Ronald Press, 1946.
3. ALEXANDER, F. & FRENCH, T. M. *Studies in psychosomatic medicine.* New York: Ronald Press, 1948.
4. AX, A. F. The physiological differentiation of fear and anger in humans. *Amer. Psychologist,* 1951, **6**, 271. (Abstract)
5. BEIER, E. G. The problem of anxiety in client-

centered therapy. *J. consult. Psychol.*, 1951, **15,** 359–62.

6. Cameron, D. E. Observations on the patterns of anxiety. *Amer. J. Psychiat.*, 1944, **101,** 36.
7. Cohen, A. R. The effects of situational structure and individual self esteem on threat-oriented reactions to power. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
8. Dibner, A. S. Cue counting: A measure of anxiety in interviews. *J. consult. Psychol.*, 1956, **20,** 475–78.
9. Dibner, A. S. The relationship between ambiguity and anxiety in a clinical interview. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
10. Frenkel-Brunswik, E. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception—an approach to personality.* New York: Ronald Press, 1951. Pp. 356–421.
11. Freud, A. *The ego and the mechanisms of defence.* New York: International Universities Press, 1946.
12. Goldstein, K. *The organism: a holistic approach to biology.* New York: American Book, 1939.
13. Haggard, E. A. Experimental studies in affective processes: Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. *J. exp. Psychol.*, 1943, **33,** 257–84.
14. Hanfmann, Eugenia. Psychological approaches to the study of anxiety. In P. Hoch and J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950. Pp. 51–70.
15. Hilgard, E. R. The role of learning in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception—an approach to personality.* New York: Ronald Press, 1951. Pp. 95–121.
16. Kruskal, W., & Wallis, W. Use of ranks in one-criterion variance analysis. *J. Amer. statis. Ass.*, 1952, **47,** 583–621.
17. Kubie, L. S. A physiological approach to the concept of anxiety. *Psychosom. Med.*, 1941, **3,** 263–76.
18. Lacey, J. I., & Van Lehn, R. Differential emphasis in somatic response to stress. *Psychosom. Med.*, 1952, **14,** 71–81.
19. Lasswell, H. D. Verbal references and physiological changes during the psychoanalytic interview: a preliminary communication. *Psychoanal. Rev.*, 1935, **22,** 10–24.
20. Luchins, A. The stimulus field in social psychology. *Psychol. Rev.*, 1950, **57,** 27–30.
21. Malmo, R. B., & Shagass, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11,** 9–24.
22. Malmo, R. B., & Shagass, C. Physiologic study of symptom mechanisms in psychiatric patients under stress. *Psychosom. Med.*, 1949, **11,** 25–29.
23. Osburn, H. G. An investigation of the ambiguity dimension of counselor behavior. Unpublished doctoral dissertation, Univer. of Michigan, 1951.
24. Postman, L., Bruner, J. S., & McGinnies, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43,** 142–54.
25. Reich, W. *Character analysis.* New York: Orgone Institute Press, 1949.
26. Rosenberg, E. Anxiety and the capacity to bear it. *Int. J. Psychoanal.*, 1949, **30,** 1–12.
27. Shire, A. Personality correlates of preferences among psychoanalytic mechanisms of defense. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
28. Sinha, A. Experimental induction of anxiety by conditions of uncertainty. Unpublished doctoral dissertation, Univer. of Michigan, 1950.
29. Sinnett, E. R. An experimental investigation of the defense preference inquiry for the Blacky Pictures. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
30. Sutich, A. The growth-experience and the growth-centered attitude. *J. Psychol.*, 1949, **28,** 293–301.
31. Van der Valk, J. M., & Groen, J. Electrical resistance of the skin during induced emotional stress. *Psychosom. Med.*, 1950, **12,** 303–14.
32. Woodworth, R. S. *Experimental psychology.* New York: Henry Holt, 1938.

Received August 29, 1956.

THE APPLICATION OF LEARNING THEORY TO THE TREATMENT OF TICS

AUBREY J. YATES[1, 2]

Institute of Psychiatry, Maudsley Hospital, London

SINCE the classic work in 1907 by Meige and Feindel (19) on the treatment of tics as examples of learned habits, little progress in this field has been reported, although their results were encouraging. In 1932, however, Dunlap (5) claimed that undesirable habits (including tics) could be eliminated by making the subject repeat them voluntarily and deliberately. But Dunlap gave few details of his procedure and no experimental demonstration of its validity, merely asserting that the tics disappeared rapidly. While his technique of negative practice has been accepted (at least in so far as the treatment of tics is concerned) in at least one modern textbook of psychology (1), an adequate rationale and experimental confirmation of its utility are still lacking.

More recently, attempts have been made to derive methods for the treatment of behavioral disorders from the constructs of modern learning theory (3, 12, 13, 16, 21, 22, 25, 26).[3] This paper reports a rationale for such treatment in the case of a tiqueur and the results of experiments carried out in order to test the validity of the deductions made from the theory.

THE TIC AS A LEARNED HABIT

The following theoretical model is proposed to show that some tics can be conceptualized as learned responses of a particular kind. The justification for proposing such a model lies in the deductions which are derived from it concerning treatment.

It is hypothesised that some[4] tics may be drive-reducing conditioned avoidance responses, originally evoked in a highly traumatic situation. In this situation, intense fear is aroused and a movement of withdrawal or aggression is made. If the movement produces or coincides with the cessation of the fear-inducing stimulus, it acquires strength through reinforcement. On subsequent occasions, through stimulus-generalization (including internal symbolization), conditioned fear ("anxiety") may be aroused, which is then reduced by the performance of the movement. In this way, the tic comes to be elicited by a large variety of stimuli and eventually achieves the status of a powerful habit.

Several points should be noticed about this model. First, although no direct experimental evidence is available to show that tics do arise in this way, two lines of investigation suggest that the formulation is a reasonable one. Wolpe (26), Solomon and Wynne (24), and others have demonstrated that animals placed in a highly traumatic situation develop conditioned avoidance responses which apparently reduce the anxiety associated with the original situation and which are highly resistant to extinction. Again, clinical investigations of the history of tiqueurs (8) have led some workers to suggest that the tic is an avoidance response arising originally in a highly traumatic situation, especially in childhood. Second, the kind of response evoked may be determined partly at least by the mode of response characteristic of the subject in any stressful situation. Malmo and his colleagues (17, 18) have presented evidence suggesting that some neurotics react to stress with cardiovascular responses, whereas others react to the same stress with muscular responses. Experiments by Lacey, Bateman, and Van Lehn (14) similarly suggest that the nature of the stressful situation is not a major variable, at least so far as autonomic functions are concerned.

The suggested model follows closely the

---

[1] Now at the University of New England, Armidale, New South Wales, Australia.

[2] The author would like to thank J. Inglis and H. G. Jones for helpful discussion on the theoretical part of the paper, and Linford Rees for permission to carry out the investigation.

[3] Earlier experiments have been reviewed by Hilgard and Marquis (9), Crafts *et al.* (4), and Lehner (15).

[4] It is not suggested that tics can arise only in the way hypothesized. Some tics appear to arise from simple imitation, expecially in young children, or as an avoidance response originally arising from some irritating but nontraumatic stimulus such as a tight collar. The proposed treatment, however, should be applicable to any tic which is a learned response.

two-factor theory of learning expounded by Mowrer (21). In terms of Hullian (10) learning theory, the reaction potential of the tic at a given moment may be conceived as a multiplicative function of the habit strength (${}_{S}H_{R}$) of the tic (determined mainly by the number of times it has previously been evoked) and the momentary drive strength of anxiety (D), which fluctuates from time to time. Since habit strength increases as a simple negatively accelerated positive growth function and eventually reaches an asymptote, further performance of the tic cannot increase its habit strength beyond a given point.

DERIVATION OF METHOD OF TREATMENT

According to the model, then, the tic may be treated as a simple learned habit which has attained its maximum habit strength. In terms of the theory, it should be possible, therefore, to extinguish the habit by building up a negative or incompatible habit of "not performing the tic." If the subject is given massed practice in the tic, then reactive inhibition (I_R) should build up rapidly.[5] When I_R reaches a certain critical point, the patient will be forced to "rest" or not perform the tic. This habit (${}_{S}I_{R}$) of not performing the tic will be associated with drive-reduction due to the dissipation of I_R and hence will be reinforced. With repeated massed practice, therefore, a negative habit ("not doing the tic") will be built up, incompatible with the positive habit of doing the tic. Furthermore, the repeated voluntary evocation of the tic should *not* serve to increase the habit strength of the tic, since it is already asymptotic and consequently not subject to strengthening by massed practice.

At any given moment, therefore, the effective reaction potential of the tic (i.e., the *ability* of the patient to perform the tic) will be a resultant of the forces in the equation

$${}_{S}\overline{E}_{R} = ({}_{S}H_{R} X D) - (I_{R} + {}_{S}I_{R})$$

As ${}_{S}I_{R}$ increases, ${}_{S}\overline{E}_{R}$ should diminish to a point at which the excitatory and inhibitory parts of the equation are equal. The response tendency of the tic will then be below threshold level, and behaviorally, extinction should be complete. During this process, of course, ${}_{S}\overline{E}_{R}$ would be expected to show fluctuations, since it is affected by a number of factors (notably drive level) which are uncontrolled (${}_{S}O_{R}$). Any increase in drive may, of course, temporarily mask the effects of an increase in ${}_{S}I_{R}$. It is interesting to note that Calvin *et al.* (2) recently produced evidence to show that the continued evocation of a habit under conditions of massed practice in rats eventually leads to its extinction, although the process may be a slow one.

The optimum conditions of massed practice under which conditioned inhibition (${}_{S}I_{R}$) may be expected to grow most rapidly and effectively are unknown with respect to the extinction of tics. The present experiments represent, therefore, an attempt to discover these conditions.

THE PATIENT

The *S* was a female psychiatric patient, 25 years old, who was highly cooperative and apparently had a strong desire to get well. She was of high average intelligence, and was markedly neurotic and slightly more extroverted than the average normal as measured by the *Guilford-Martin Personality Inventory of Factors S, T, D, C, R.*[6] Four clear-cut tics were chosen for the experiment: a complex stomach-contraction breathing tic, a nasal "explosion" (expiration), a coughing tic, and an eye-blink tic. These tics appeared to have started originally following two very traumatic experiences about ten years previously when *S* felt that she was being suffocated while undergoing anaesthesia; she said she was terrified that she was going to die and struggled madly. She could not bear the thought of an anaesthetic mask and could not tolerate any object's being placed over her face. Superficially, at least, these tics seemed to be conditioned avoidance responses originally established in a traumatic situation.[7] She also said she felt a need to do the tics and experienced relief when they occurred (i.e., they were drive-reducing). They varied mark-

---

[5] With massed practice the tic will also, for the most part, cease to be reinforced, because, since anxiety is only intermittent in *S*, it will be unaccompanied by drive reduction.

[6] Eysenck (7) has shown that these scales are reasonably good measures of neuroticism (C and D scales) and introversion-extraversion (R scale).

[7] Presumably, escape in such a situation from the noxious stimulation was effected by the patient's becoming unconscious.

TABLE 1

CHANGES IN MEAN FREQUENCY OF FOUR TICS UNDER CONDITIONS OF VOLUNTARY EVOCATION

Stomach			Eyeblink			Nose			Throat		
Sessions	M	σ	Sessions	M	σ	Sessions	M	σ	Sessions	M	σ
1–50	52.83	4.06	1–50	36.96	4.74	1–50	45.56	5.41	1–50	42.00	4.17
51–100	40.54	3.90	51–100	31.96	2.48	51–100	35.11	3.02	51–100	32.18	3.55
112–150	41.69	2.58	101–150	31.30	1.63	112–150	32.06	3.13	101–150	31.52	3.53
151–175	19.40	2.15	151–200	26.87	2.00	151–180	23.71	2.73	151–200	23.91	2.25
176–200	15.26	1.57									
201–220	11.95	1.09	201–220	26.04	1.60	201–220	18.97	4.25	201–220	20.37	3.33
221–285	9.67	1.04	221–285	20.34	1.74	221–248	13.23	2.12	221–285	15.51	2.10
286–315	8.11	0.75	286–315	21.14	1.48	286–315	8.05	1.90	286–315	12.88	1.73

edly from time to time in frequency of occurrence. It was decided to treat the four tics concurrently but independently.

EXPERIMENT I

Sessions 1–100

The purpose of this experiment was to test the general hypothesis that massed practice leads to a significant decrement in the ability of *S* to respond voluntarily.

Two sessions per day were carried out, each consisting of five one-minute trials for each tic under conditions of massed practice, with one minute's rest between each set of five trials.[8] *S* was instructed to reproduce each tic as accurately as possible and to repeat it without pause during the practice periods. No stress was laid on speed. A complete session lasted exactly 45 minutes. One such session was carried out at the hospital each day under supervision; *S* carried out the other on her own. The order of practising the tics was varied at random. The score recorded was the number of tics per minute as counted by *E* in the case of the nasal and throat tics and by *S* in the case of the stomach and eyeblink tics. This procedure was followed for 100 sessions (50 days).

The frequency of occurrence of each tic per minute (under test conditions) almost invariably showed a steady decline from trials one to five within a single session. It was considered that the average of the five trials per session would give a reliable estimate of the strength of the tic under conditions of voluntary evocation in any particular session. The mean score for the first 50 sessions was compared with the mean score for the second 50 sessions for each tic separately.

The results (Table 1) are remarkably consistent for the four tics. The mean score for the second set of 50 sessions is significantly lower than the mean for the first set of 50 for each tic (values of t were as follows: stomach, 15.42; eyeblink, 6.60; nose, 11.93; throat, 12.68; p was $< .001$ in each case). The results, therefore, are in accordance with prediction and offer support for the general theory.

EXPERIMENT II

Sessions 101–150

Three hypotheses were tested over Sessions 101–150.

Hypothesis I. It seemed reasonable to contend that further massed practice under standard conditions would produce a further decline in the tic frequency. To test this hypothesis, the eyeblink and throat tics were practised under standard conditions.

Hypothesis II. The difference between means of the first and second sets of 50 sessions in Experiment I might be due to I_R not completely dissipating between sessions. To test this hypothesis, practice on the stomach tic was discontinued for 11 sessions; during the ensuing Sessions 112–150, standard conditions were resumed. If the decline in Experiment I were due to I_R alone, the mean for Sessions 112–150 should not be lower than that for Sessions 1–50, i.e., recovery of function should take place.

Hypothesis III. To test the notion that the growth of $_sI_R$ would be facilitated by increasing the duration of massed practice, the nasal tic was practised for 15 min. daily without a break under instructions to keep the amplitude as strong and as constant as possible. This procedure apparently had the desired effect of increasing the number of involuntary rest pauses, a crucial factor in the growth of $_sI_R$, since in the last few minutes of the period, the rate of responding typically dropped to 3–4 per minute. Five sessions (one per day) were carried out in this way, the prolonged period always following practice on the other tics to avoid interference with them. No standard practice on the nasal tic was carried out at this time. During Sessions 112–150, standard conditions were imposed. It was predicted that the mean for Sessions 112–150 would be significantly below that for Sessions 51–100.

The results of this experiment are reported in Table 1. Hypothesis I was not confirmed. For both eyeblink and throat tics, the mean frequency per session was not significantly below that for Sessions 51–100; (for the eyeblink, $t = 1.570$, $p = > .05$; for the throat, $t = 0.927$, $p = > .05$). However, no recovery of function took place, since in both cases the mean for Sessions 101–150 was significantly lower than that for Sessions 1–50; (for the eyeblink, $t = 7.982$, $p = < .001$; for the throat, $t = 13.563$, $p = < .001$).

Hypothesis II was not confirmed. The mean for Sessions 112–150 was significantly lower than the mean for Sessions 1–50 ($t = 14.928$, $p = < .001$). However, the mean for Sessions 112–150 was not significantly different from the mean for Sessions 51–100, ($t = 1.589$, $p = > .05$).

Hypothesis III was confirmed. The mean for Sessions 112–150 was significantly lower than the mean for Sessions 51–100, ($t = 4.655$, $p = < .001$). Although the decline was not very large, it should be remembered that only five sessions of increased massed practice were given.

---

[8] Designated henceforward the standard procedure.

During this experiment, therefore, *S* seemed to be on a plateau with respect to three of the tics, but recovery of function did not take place, even when practice was entirely discontinued for a week. Further, it was apparent that the growth of $_{s}I_{R}$ could be facilitated in some degree by increasing the amount of massed practice.

EXPERIMENT III

Sessions 151–200

Three hypotheses were again tested over Sessions 151–200.

Hypothesis I. Practice of the eyeblink and throat tics was continued under standard conditions on the assumption that during Sessions 50–150 *S* had been on a plateau with respect to these tics; it was predicted that further practice would eventually induce a decline.

Hypothesis II. It seemed probable that within the standard practice conditions, the intensification of the response would increase the rate of decline. Dunlap (5) stressed the importance of instructing *S* to pay careful attention to the tic. *S* was, therefore, asked to reproduce the stomach tic as precisely and as intensively as possible. The conditions of practice were otherwise standard.

Hypothesis III. The prolongation of the period of practice for the nasal tic was continued. However, the method was varied. Each day, at the *end* of the hospital session only, *S* did 15 min. massed practice of the nasal tic as in Experiment II. In the second session of that day, the nasal tic was not practised (since the I_R associated with the 15-min. practice period might not have wholly dissipated and could affect performance). However, at the *beginning* of the hospital session each day, the nasal tic was practised under standard conditions. This served as a measure of the influence of the previous day's intensive practice. The standard nasal practice was followed by the throat and eyeblink tics, which always preceded the stomach tic to avoid the effects of fatigue induced by that tic under intensive conditions. In this way, at least 12 hours elapsed between the intensive and the standard nasal practises. During this experiment, there were 20 intensive and 30 standard practice periods for the nasal tic.[9]

The results, as shown in Table 1, supported Hypothesis I. The mean for Sessions 151–200 was significantly lower than the mean for Sessions 101–150 for both eyeblink and throat tics, (for the eyeblink, $t = 12.137$, $p = < .001$; for the throat, $t = 9.931$, $p = < .001$).

With regard to Hypothesis II, the results obtained with the stomach tic were not directly comparable with those obtained in Sessions 1–150 because of the altered instructions within the standard conditions. When, however, Sessions 151–175 were compared with Sessions 176–200, a significant decline was apparent in the latter set, ($t = 7.793$, $p = < .001$). The hypothesis may, therefore, be regarded as supported.

Hypothesis III was also supported. Extensive massed practice led to a striking decline in frequency, the mean for Sessions 151–180 being significantly below that for Sessions 112–150 ($t = 11.608$, $p = < .001$).

The results of Experiment III suggested, therefore, that the standard procedure led to a continuing decline in frequency, that extensive massed practice had a very striking effect on frequency of response, and that increasing the intensity of response within the standard situation also led to a decline in frequency.

EXPERIMENT IV

Sessions 201–220

The purpose of Experiment IV was to test more rigorously the hypothesis that the decline in frequency of the four tics was a function of the growth of $_{s}I_{R}$ rather than of the accumulation of I_R. When practice on the stomach tic was discontinued in Experiment II, *S* later reported that she had been unable to inhibit it during the 15-min. practice of the nasal tic. Hence, the earlier experiment seemed inconclusive. *S* was, therefore, given a two-week holiday during which no practice of any kind was carried out. It was predicted that no recovery of function would take place. Following this interlude, the four tics were practised under standard conditions only, for 20 sessions.

The results, as shown in Table 1, indicated that the eyeblink tic showed no change following complete rest, ($t = 1.00$, $p = > .05$). The other three tics all showed a significant further decline during Sessions 201–220 compared with the level reached in Experiment III (for the stomach, $t = 8.37$, $p = < .001$; for the nose, $t = 3.98$, $p = < .001$; for the throat, $t = 4.49$, $p = < .001$). The decline was particularly striking in the case of the nasal tic. The hypothesis that no recovery of function would occur seemed, therefore, to be supported, but the significant decline following rest in the case of three tics remained to be explained.

EXPERIMENT V

Sessions 221–275

Three hypotheses were tested in this experiment.

Hypothesis I. Massed practice under standard conditions was continued for the eyeblink and throat tics, the prediction being that a further significant decline would result.

Hypothesis II. Massed practice under standard conditions but with intensification of the response was continued for the stomach tic, the prediction being that a further significant decline would result.

Hypothesis III. It was decided to test whether the rate of growth of $_{s}I_{R}$ could be significantly increased by an increase in the *rate* and *intensity* of repetition of the nasal tic. To test this hypothesis, the 15-min. period of massed practice was used. This period was divided into periods of 5 seconds of massed practice followed by 10 seconds of rest. During the 5 seconds, *S* performed the nasal tic as rapidly and as intensively as possible. It was expected that I_R would accumulate more rapidly and would dissipate less between indi-

---

[9] The numbers are unequal because *S* was later allowed to do the 15 min. practice on her own on weekends.

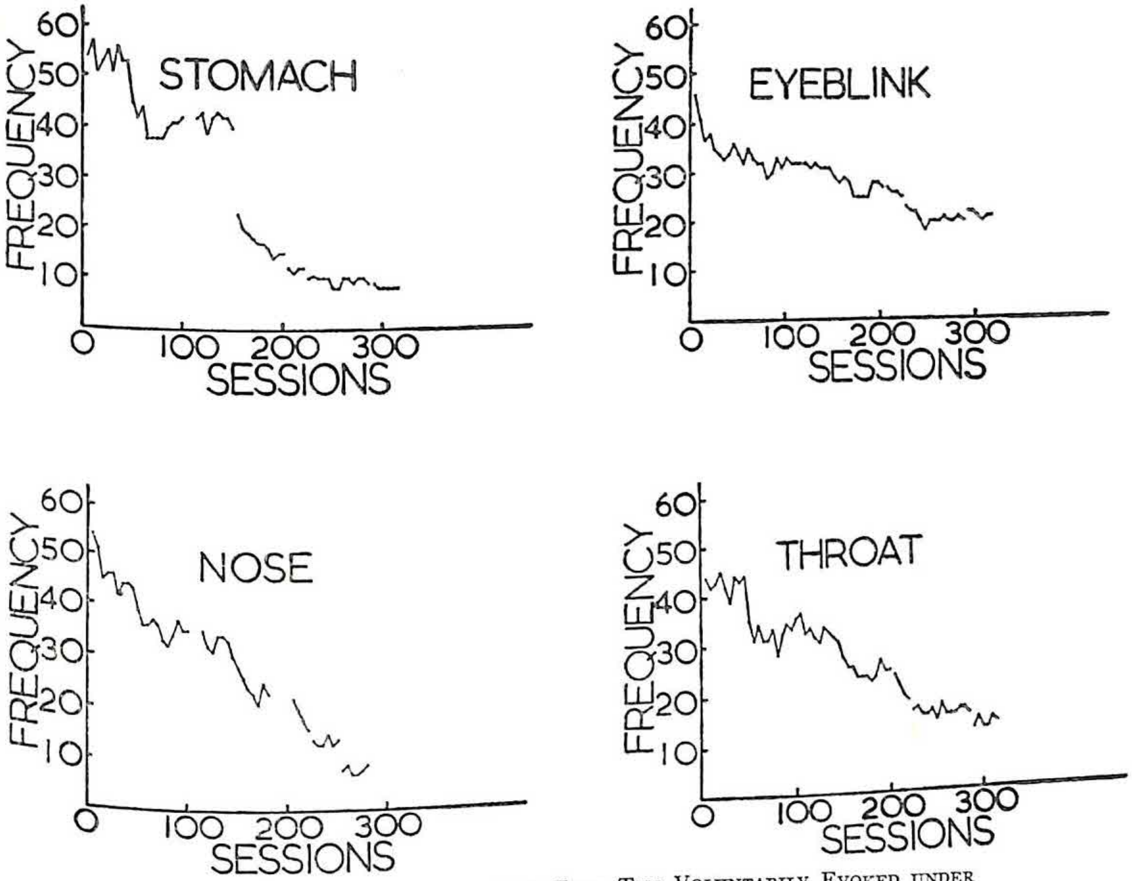


FIG. 1. CHANGES IN FREQUENCY OF FOUR TICS VOLUNTARILY EVOKED UNDER VARIOUS CONDITIONS OF MASSED PRACTICE

vidual tics. Because 10 seconds of rest would thus be expected to have a high reward value, $_sI_R$ should be more strongly reinforced. Apart from the change in procedure for the 15 minutes of practice, the method adopted was identical with that described in Experiment III.

The results, as shown in Table 1, indicated that all three hypotheses were confirmed[10] (for the eyeblink, $t = 13.67$; for the stomach, $t = 8.26$; for the nose, $t = 5.21$; for the throat, $t = 6.16$; p was $< .001$ in each case).

EXPERIMENT VI

Sessions 276–315

The purpose of this experiment was to test the hypothesis that a very prolonged period of massed practice followed by a prolonged rest would lead to a significant decrement in response. It had been found that in Sessions 201–220 a highly significant decline took place on three of the four tics following two weeks' rest, the decline being most striking in the case of the nasal tic. The possibility arose, therefore, that the crucial set of conditions for increasing $_sI_R$ rapidly lay in a combination of very intensive practice followed by a long rest. In general, it was hypothesized that optimum increments of $_sI_R$ would be obtained if sufficient massed practice were given to allow the accumulation of a maximum amount of I_R, followed by a rest period sufficient to allow the complete dissipation of I_R. To test this hypothesis, four one-hour sessions were given, during which uninterrupted massed practice of the nasal tic was carried out. The other tics were not practised during these sessions, each of which was separated from the next by a three-day interval. There were no standard practices of the nasal tic during this period. The other tics were practised under the usual conditions during the second daily session only (i.e., once per day, constituting Trials 276–285). At the end of this period, the patient was given a three-week holiday during which no practice of any kind was carried out. Following this rest, 30 standard sessions (286–315) were given for all four tics.

The results, as shown in Table 1, indicate that the nasal tic showed a very significant decline in the mean for Sessions 286–315 compared with Sessions 221–248 immediately preceding the four one-hour practices ($t = 7.99$, $p = < .001$). Throat and stomach tics showed a significant but much smaller decline (for the throat, $t = 6.43$, $p = < .001$; for the stomach,

---

[10] Table 1 shows the means for Sessions 221–285 for eyeblink, throat, and stomach tics. Sessions 276–285 were, however, actually performed for these tics during Experiment VI while the nasal tic was being given four one-hour practices.

$t = 8.32$, $p = < .001$). The eyeblink tic showed a small but significant recovery ($t = 2.32$, $p = .02$). Thus, the hypothesis was generally supported.

At the conclusion of this experiment, the mean score for the last 30 sessions (286–315) was compared with the mean score for the first 30 sessions (1–30).[11] The percentage decrement in the frequency of voluntary responding for each tic was as follows: nasal, 84.14; throat, 68.83; stomach, 63.42; eyeblink, 51.47. The course of decline is shown in Fig. 1, in which each point represents the average of five sessions (i.e., 25 one-min. trials). Gaps in the curves indicate rests or changes in treatment.

CHANGES IN INVOLUNTARY TICS

It is notoriously difficult to obtain a reliable estimate of the rate of involuntary responding in tics, which fluctuates markedly from day to day, even from hour to hour. While a *change* in the rate of response would be very difficult to assess reliably, a complete cessation of involuntary responding outside the test situation would be apparent to *S*, to her acquaintances, and to *E*. *S* had not reached this clinical criterion at the end of Experiment VI. At the end of the series, however, she was asked to write a frank account of her experience over the period of treatment, both in relation to the tics and to her general adjustment.

Her report indicated that the nasal tic "has almost vanished. It returns occasionally . . . every now and then . . . but most definitely it does not plague me every few minutes as before. It is absent for days on end." The eyeblink and throat tics were reported as sharply reduced in frequency. The stomach tic, however, had changed very little. She also reported that she had "improved in all departments generally," was "able to relax in a bus or train, able to read or execute a crossword puzzle." There was continual reference to the fact that not doing the movement did not leave her frustrated. She concluded that she was, "not tired of the treatment, if anything now that I have a taste of success the desire is greater." She also pointed out that "one of the striking features is that my closest friends notice the change." The specific and general changes were confirmed by her physician.[12]

DISCUSSION

The results obtained in these experiments support the theory proposed to explain the genesis of certain tics and the deductions made from it regarding a rational method of treatment. Within the test situation there was objective evidence of a striking decline in the rate of voluntary responding for all tics. The rate of decline was most striking in that tic (nasal) which received the largest amount of massed practice. There was no evidence of any recovery of function following prolonged rest. The data further suggest that the optimum conditions for the production of conditioned inhibition may be a period of very prolonged massed practice, leading to the accumulation of large amounts of reactive inhibition, followed by a very prolonged rest to allow for the total dissipation of reactive inhibition. There was subjective evidence suggesting that conditioned inhibition generalized beyond the test situation itself and that most generalization occurred with respect to the tic which had received most massed practice. It is not suggested, however, that the results outlined above can be unequivocally interpreted. Under clinical conditions of testing and with limited time available, perfect experimental control and balance of design proved impossible to obtain. Thus, it is possible that the decline in Sessions 286–315 in the nasal tic was a function entirely of the amount of massed practice given and not of the combination of practice and rest. Although no standard practices were carried out on the nasal tic between the four one-hour periods, *S* reported that she became much worse outside the test situation during this time. However, during the three weeks' rest period which followed, she reported that she rapidly lost the tendency to perform the tic involuntarily, which suggests that the rest was an essential part of the treatment. These results imply that the cumulative reinforcement associated with the dissipation of reactive inhibition may continue for longer periods of time than is commonly supposed. This interpretation is consistent with the findings of Jahnke and Duncan (11).

A number of considerations which arose during the experiments merit brief mention. It would have been useful to have left one tic untreated and to have taken control readings under standard conditions at regular intervals. Under these conditions, little or no conditioned inhibition would be expected and the prediction would be that the untreated tic would not decline during the course of the experiment. Second, an objective measure of the frequency of each tic outside the test situation at the beginning and end of the experiments is highly desirable. It should not be too difficult to measure the frequency of tics reliably in a natural situation. Third, attention should be paid to the amplitude as well as the frequency of the tics; a measure of both variables would be useful. Fourth, *S* should be instructed from the start to reproduce the tic as accurately as possible in order to facilitate the growth of I_R.

---

[11] For the stomach tic, the comparison was between Sessions 151–180 and 286–315.

[12] During these experiments, *S* received no other treatment but saw her physician for a few minutes about once a month.

Nevertheless, certain features of the obtained results are highly suggestive. The shape of the curves suggests, in some respects at least, the obverse of a learning curve. There is no evidence of an initial rise in frequency, indicating that the habit strength of the tic was probably at its asymptote before massed practice was begun. Again, the decline in frequency appears to proceed according to the usual learning process. There are periods of stability (plateaus) followed by sudden falls, suggesting that a negative habit was actually being learned. This finding agrees with the results of Edmonson and Amsel (6), who found that in the extinction of conditioned avoidance responses, long periods during which no change took place, alternated with sudden sharp declines. The curves also show the sessional fluctuations common in learning curves, where all the relevant factors (notably, in this instance, drive level) cannot be controlled.

An important problem concerns the relationship between the anxiety and the tic. *S* reported at the end of treatment that she was much less anxious than formerly (reflected in many particular examples of behavior she gave). This reduction in anxiety follows directly from the theory. If the ability to respond is significantly reduced, then the anxiety leading to the tics will persist. However, since no traumatic event happens in fact as a result of the anxiety, the latter, being itself a conditioned response, gradually extinguishes. This formulation suggests an alternative method of treatment. If the tics are prevented from occurring, then reduction of the anxiety should follow. When this happens, the tics will also decline, since they are a response to anxiety. Meige and Feindel (19) reported cases in which paralysis of facial tics for several days led to their disappearance for a period of several months. This hypothesis also receives support from the findings of Solomon and his colleagues (23) that conditioned avoidance responses can be extinguished in dogs by presenting the CS a large number of times but preventing the avoidance response by restraint. The usual extinction procedures, which formerly had been unsuccessful, could then be applied. These arguments are the principal justification for symptomatic treatment. It is interesting to note that they follow directly from Mowrer's two-factor theory. Mowrer himself has strongly opposed symptomatic treatment (21, p. 620) although making use of it in the treatment of enuresis. In terms of the theory, it would also be predicted that this method of treatment would be most successful with children because ${}_SH_R$ would presumably be at less than asymptotic strength. Thus, as soon as negative practice began, most of the repetitions of the tic would be contributing towards the growth of ${}_SI_R$ rather than ${}_SH_R$, and the growth of the negative habit (not-doing-the-tic) would proceed more rapidly than the growth of the positive habit.

SUMMARY

A theoretical model was proposed to show that some tics may be conceptualized as drive-reducing conditioned avoidance responses, originally evoked in a traumatic situation.

From this model, a method of treatment was derived. It was predicted that if the tics were evoked voluntarily under conditions of massed practice, a negative habit of "not doing the tic" should be built up, resulting ultimately in the extinction of the tics and that this extinction should generalize beyond the test situation.

The results of a number of experiments support the validity of the theory. There was a significant decline in the ability to respond voluntarily under various conditions of massed practice and rest.

The optimum condition for the growth of

the negative habit appeared to be the combination of very prolonged massed practice followed by prolonged rest.

Subjective reports of the patient indicated considerable clinical improvement outside the immediate test situation.

REFERENCES

1. BORING, E. G., LANGFELD, H. S., & WELD, H. P. *Foundations of psychology.* New York: John Wiley, 1948.
2. CALVIN, A. D., CLIFFORD, L. T., CLIFFORD, B., BOLDEN, L., & HARVEY, J. Experimental validation of conditioned inhibition. *Psychol. Rep.*, 1956, **2**, 51–56.
3. CHERRY, C., & SAYERS, B. MCA. Experiments upon the total inhibition of stammering by external control and some clinical results. *J. psychosom. Res.*, 1956, **1**, 233–246.
4. CRAFTS, L. W., SCHNEIRLA, T. C., ROBINSON, ELSA E., & GILBERT, R. W. (Eds.). *Recent experiments in psychology.* New York: McGraw Hill, 1950.
5. DUNLAP, K. *Habits: Their making and unmaking.* New York: Liveright, 1932.
6. EDMONSON, B. W., & AMSEL, A. The effects of massing and distribution of extinction trials on the persistence of a fear-motivated instrumental response. *J. comp. physiol. Psychol.*, 1954, **47**, 117–123.
7. EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. di Psicol.*, 1957, **50**, 113–140.
8. GERARD, MARGARET. The psychogenic tic in ego development. In Anna Freud, *et al.* (Eds). *Psychoanalytic study of the child.* Vol. 2. New York: Inter. Univer. Press, 1946, Pp. 133-162.
9. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
10. HULL, C. L. *Principles of behavior.* New York: Appleton Century, 1943.
11. JAHNKE, J. C., & DUNCAN, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, **52**, 273–282.
12. JONES, E. L. Exploration of experimental extinction and spontaneous recovery in stuttering. In W. Johnson (Ed.), *Stuttering in children and adults.* Minneapolis: Univer. of Minnesota Press, 1955. Pp. 226–231.
13. JONES, H. G. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 414–419.
14. LACEY, J. I., BATEMAN, D. E., & VAN LEHN, R. Autonomic response specificity. *Psychosom. Med.*, 1953, **15**, 8–21.
15. LEHNER, G. F. J. Negative practice as a psychotherapeutic technique. *J. gen. Psychol.*, 1954, **51**, 69–82.
16. LIVERSEDGE, L. A., & SYLVESTER, J D. Conditioning techniques in the treatment of writer's cramp. *Lancet*, 1955, **2**, 1147–1149.
17. MALMO, R. B., & SHAGASS, C. Physiologic study of symptom mechanisms in psychiatric patients under stress. *Psychosom. Med.*, 1949, **11**, 25–29.
18. MALMO, R. B., SHAGASS, C., & DAVIES, F. H. Symptom specificity and bodily reactions during psychiatric interview. *Psychosom. Med.*, 1950, **12**, 362–367.
19. MEIGE, H., & FEINDEL, E. *Tics and their treatment.* London: S. Appleton, 1907.
20. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*, New York: John Wiley, 1951. Pp. 435–472.
21. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.
22. RAYMOND, M. J. Case of fetishism treated by aversion therapy. *Brit. med. J.*, 1956, **2**, 854–857.
23. SOLOMON, R. L., SIDD, J. J., WATSON, P. D., & BLACK, A. H. The use of d-tubocurarine in the extinction of fear in dogs. *Amer. Psychologist*, 1955, **10**, 395. (Abstract)
24. SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: the principles of anxiety conservation and partial irreversibility. *Psychol. Rev.*, 1954, **61**, 353–385.
25. WISCHNER, G. J. Stuttering behavior and learning: A preliminary theoretical formulation. *J. Speech Dis.*, 1950, **15**, 324–335.
26. WOLPE, J. Experimental neuroses as learned behavior. *Brit. J. Psychol.*, 1952, **43**, 243–268.

Received October 10, 1956.

GSR CONDITIONING IN ANXIETY STATES, NORMALS, AND CHRONIC FUNCTIONAL SCHIZOPHRENIC SUBJECTS[1]

EDMUND S. HOWE[2]

University College, London

In recent years, increasing attention has been given to experimental studies of anxiety and conditioning and to interpathological explorations of conditioning. There is now general agreement that anxiety, assessed either psychiatrically or psychometrically, plays a part in human eyelid (22, 23, 24, 25) and GSR conditioning (2, 26, 27). Several studies have now confirmed that the neurotic *S* conditions more readily in a threat situation than does the normal (23, 25), a difference which may be explained by an increase in motivation derived from anxiety. The data on the conditionability of psychotics are, on the other hand, rather scanty and unsystematic (1, 15, 19, 21). Two recent studies (23, 25) have reported data and deductions for psychotics with the conclusion that these *S*s condition more readily than both neurotic and normal *S*s. This conclusion is not consistent with the findings of Paintal (18), for example, who examined GSR to electric shock and to threat of shock in 450 normal and 450 psychotic *S*s. Although there was little difference between the *S*s' response to the shock, there was a significantly smaller response in psychotic *S*s to the *threat* of shock. The present study examined conditioned GSR behavior in hospitalized anxiety states, normals, and hospitalized chronic functional schizophrenics, on the hypothesis that these three groups involve, respectively, a descending order of total drive strength (Hull's *D*) with respect to such a nociceptive stimulus as electric shock.[3]

The question of differential drive strengths in the three subject groups may, of course, be approached empirically by a comparison of conditioning measures, but this procedure is not necessarily the optimal one. From the standpoint of Hullian theory (11), provision is made for a multiplicative combination of other drives (e.g., anxiety) with the relevant drive energizing the habit involved. In the present case, *D* may therefore be considered a direct function of anxiety level. In the case of the anxiety states, this contention has repeatedly proved fruitful (22, 23, 24, 25); i.e., there is a sound basis for predicting that the anxiety cases will show more rapid, stronger conditioning than normal controls. There is, however, a difficulty of relating Hull's *D* to the anxiety level of schizophrenics because, while such patients may well show an absence of *manifest* anxiety symptoms (leading to a prediction of slower conditioning), it may be reasonably contended (e.g., 23, 25) that their withdrawal is a defense against high (hypothetical) *internal* anxiety (leading to a prediction of more rapid conditioning). Perhaps this problem has been turned into an either/or question, neglecting the possibility that the *degree* of chronicity and deterioration in schizophrenia at least partly determines anxiety level and hence reactivity in a threat situation, the relationship probably being an inverse one.

---

[1] This paper contains a portion of the data from a thesis submitted to The Department of Psychology, University College, in candidacy for the degree of doctor of philosophy in the University of London. The writer is indebted especially to John W. Whitfield, who throughout supervised the research, for his encouragement and advice; and to Roger W. Russell and Hans J. Eysenck, each of whom made valuable comments at different stages of the work. Many thanks are also due to the members of the psychiatric, psychological, and nursing staffs at St. Ebbas Hospital, Epsom, and The Belmont Hospital, Sutton, Surrey, for their willing and continuous cooperation in making it possible for the writer to execute this study.

[2] Now with the Department of Psychiatry, University of Maryland School of Medicine.

[3] A second but major purpose of the study was to examine the effect of drive level *D* upon the steepness of the gradient of stimulus generalization, a relationship that has been explored by more recent writers (4, 20). This is the essential reason for extinguishing one-third of each group to a different tone. When gradients of generalization over three test points were examined at each extinction trial for each group, however, only two gradients showed a significant departure from zero slope, which could, of course, occur by chance. Because of this result, data and discussion on generalization have been entirely omitted from this report. It was reasoned *ex post factum* that the outcome concerning the gradients was a function mostly of *S*s having been given minimal familiarity with stimuli to be employed during the pretesting and conditioning phases (which would *decrease* responsiveness to stimulus differences and hence spuriously *increase* generalization). It was thus not possible to offer conclusive evidence about the hypothesis set up: namely, that amount of generalization is a function of drive strength, rather than simply of stimulus similarity (see, for example, 11, 16, 17, 20).

It can thus be argued, as Malmo *et al.* concluded, that, "the much lower reactivity of the chronic schizophrenic is evidence for a trend toward lowered reaction to pain as schizophrenia progresses toward chronicity" (14, p. 63). Malmo and his collaborators suggested, however, that this interpretation is probably applicable to rather specific experiences like pain, and may not be true at all for other situations involving overt adjustments. The characteristics of chronic schizophrenics, as presented by such writers as Hoagland (6, 7, 8), Hoskins (9), Hill *et al.* (5), and others, show a lowered reactivity to equilibrium functions, central disturbances rather than peripheral ones, and inadequate responses to stress as measured by lymphocyte counts and 17-ketosteroid output. It has also been suggested by Hoskins (9) that an inadequate adrenal cortical response, rather than a higher stress threshold, is responsible for the lowered reactivity of the chronic schizophrenic. The general picture emerging from a survey of this kind of work leads to the following conceptualization of the chronic deteriorated schizophrenic: stress-produced responses which seem to be justified biologically frequently do not occur at all, or occur late, or occur to a diminished degree. This general statement is, of course, quite the converse from and at the same time complementary to what is expected in the anxiety neurotic, despite the fact that different pain thresholds have not been demonstrated for the different *S*s. With such considerations in mind, it was considered most reasonable to hypothesize low drive strength with respect to electric shock in the schizophrenic population used in this study, and to predict weaker, slower conditioning.

METHOD

Subjects

A total of 60 normal control *S*s was used, half drawn from University of London students and half from the nursing staffs of the two cooperating hospitals. Student *S*s ranged from 19 to 24 years; half of them were male, sixty per cent taking honors in psychology and the remainder coming from other departments. The sample of nurses contained an equal number of males and females, with an age range from about 18 to 25. All control *S*s were volunteers.

The anxiety *S*s were in-patients drawn from Belmont Hospital, Surrey, a National Rehabilitation Center. A total of 60 *S*s, psychiatrically diagnosed as acute or chronic cases of severe anxiety, were distributed as follows: anxiety reactions, 50; phobics, 5; compulsives, 3; hysterics, 2. There were 29 males and 31 females. All *S*s reported upon cooperated voluntarily. The ages of these patients were not recorded, but the mean was about 30 with a range of approximately ten years.

The schizophrenic group was drawn from voluntary in-patients at St. Ebbas Hospital, Surrey, and was selected on the basis of psychiatric diagnosis. A criterion of chronicity was set at two years or more of illness; but in fact the mean duration of illness was over five years. After elimination of unusable data (due, for example, to some *S*s' quitting midway through the conditioning procedure), 60 *S*s psychiatrically diagnosed as chronic, functional schizophrenic were distributed as follows: simple schizophrenia, 3; paranoid schizophrenia, 4; catatonia, 10; hebephrenia, 1; and "mixed schizophrenic symptoms," 42. There were 26 males and 34 females, with a mean age of 32 years. Patients were assigned for conditioning by the psychiatric staff. Any *S* who attempted to leave the experimental room was permitted to do so.

A further group of 36 schizophrenic *S*s proved to be unconditionable according to one of three criteria: no GSR whatsoever to twice the standard intensity of shock usually delivered to other *S*s during pretesting; complete sensory adaptation to shock (i.e., no GSR) either before or at the last conditioning trial; or zero GSR to the first extinction trial. A comparison of these *S*s with those who were conditioned is presented in the results.

Apparatus and Method of Recording GSRs

The Féré method, involving impressed current, was used in the circuit described by Lacey and Siegel (13). GSRs were read from a Pye "Scalamp" galvanometer, the period of which is two seconds, and which can be read quite accurately to 0.1 mm. In order to determine *S*'s resistance, it was simply necessary to adjust R_2 until approximately 60 microamperes were passing through him. Since the potential drop across *S* was always constant, basal resistance was obtainable from a specially calibrated dial, superimposed upon R_2. The ohmic magnitude of any single GSR was thus directly proportional to the magnitude of the spot deflection. The GSR electrodes consisted of two highly polished brass discs, one inch in diameter, each firmly embedded in the base of a vacuum tube ground down to a depth of about one-half inch. Felt plugs, soaked in two per cent saline before use with each *S*, were wedged in the cups on top of the electrodes. The electrodes were firmly attached with elastic bandages to the third phalanges of the second and third fingers of each hand. It was evident from preliminary explorations that in this position, rather than in the center of the palms, the contact pressure (and hence the amount of current permitted to flow) was relatively constant.

Electric shocks were delivered from a coil built into an electronic timing mechanism. The intensity was varied with an ordinary tuning dial graduated in arbitrary units. The shock was of 75 msec. duration and was delivered across the *S*'s wrists through two polished brass electrodes.

Conditioned and other tonal stimuli were produced by a type H.1. Advance audiofrequency generator, delivered to *S* through an Erikson low impedance earphone to the preferred ear, the headset being balanced by a nonfunctional earpiece. The frequencies employed were 1967, 1000 and 468 c.p.s. Since the work of Hovland (10) has provoked considerable discussion, three of his frequencies were employed, but no check was made upon the accuracy of his psychophysical measurements. All tonal stimuli were presented at an intensity of approximately 40 decibels, but were equated for loudness in preliminary studies.

A specially built electronic device served to control the presentation of all stimuli at the touch of a post office relay switch. Tonal stimuli lasted for exactly 400 msec., followed during conditioning by an interval of 100 msec., after which the shock of 75 msec. duration was presented.

Pretesting Procedure

In all three groups, *S* was seated in one cubicle, while *E*, after installing and instructing him, controlled the whole conditioning procedure remotely from a separate cubicle. Each *S* was seated in a reclining chair with a footrest. He was first asked to relax for about three to five minutes during which period the basal resistance level typically assumed some steady point. The resistor R_2 in the galvanic circuit was then adjusted until a predetermined current of 59.3 microamps was passing through him, thus permitting measurement of basal resistance level to the nearest 1000 ohms. This initial adjustment, of course, placed a certain potential across *S*, which remained constant for him during the entire experimental session.

Pretesting of stimuli involved presentation of the 1967 c.p.s. tone once per minute for three minutes, the GSR being observed on each occasion. It was evident in the vast majority of cases that adaptation was then complete, since, at the fourth pretesting presentation of the tone, the response was either zero or negligible. No pretests of responsiveness to the other two tones were made, it being assumed that generalization of adaptation would be considerable if not complete. Finally, each *S* was given three shocks of increasing intensity and was asked on each occasion for a verbal report of how each felt (unpleasant but not painful, painful, etc.). Successful attempts were made to deliver the same mean intensity of shock to each group of *S*s.

Conditioning Procedure

After a pause of five minutes, each *S* was given eight reinforced trials of the 1967 c.p.s. tone (CS) followed by the shock (UCS) at a staggered rate of one per minute. For each stimulation, minimal and maximal galvanometric readings were taken. One minute after the eighth reinforcement trial, each *S* was subjected to experimental extinction over eight trials. For the latter purpose, one-third of each group (i.e., normal control, anxiety, and schizophrenic groups) was presented with the 1967 c.p.s. tone (*0*), another third with the 1000 c.p.s. tone (*1*), and the remainder with the 468 c.p.s. tone (*2*). Initial and maximal galvanometric deflections were noted through the eighth extinction trial.

Treatment of Data

Haggard (3) demonstrated that logarithm of conductance change is a favorable transformation of GSR data, since it has been found to satisfy the requirements of normality and independence of response magnitude from basal resistance level. Initial examination of data at the first acquisition trial indicated that this measure was superior to both resistance change and conductance change according to these criteria.

RESULTS

As a matter of course, the basal resistance levels at the beginning of the experiments were compared on an interpathological basis. Mean values in thousands of ohms and variabilities are shown in Table 1. Using separate estimates of the population variance for each group of observations, the appropriate *t* values showed that the three means differed from each other at less than the .001 level of significance. Further, the variance of the anxiety group was smaller than that of the normal group ($F = 1.58$, $p < .05$), while that of the schizophrenic group was considerably *larger* than that of the normal group ($F = 3.56$, $p < .001$). These figures point up the relatively high variability of schizophrenic patients and, indirectly, the instability of psycho-physiological systems established in such *S*s and reported by others (for example, 9).

A chi-square test of normality of response magnitudes at the first acquisition trial showed that with log conductance transformations the anxiety and normal groups could be considered as showing only random departures from normality (with 5 *df*, $\chi^2 = 8.21$ and $p = .15$ for the anxiety group; with 4 *df*, $\chi^2 = 6.77$ and $p = .15$ for the normal group), while with 5 *df*, the schizophrenic group showed a χ^2 of 12.5. The latter figure would be expected

TABLE 1

BASAL RESISTANCE LEVELS BEFORE CONDITIONING IN OHMS × 10^3

	Group		
	Anxiety	Normal	Schizophrenic
Mean	21.6	34.1	53.0
SD	3.6	5.7	20.3

to occur only 3 per cent of the time in repeated sampling and hence indicates a significant departure from normality. The first acquisition trial data were also subjected to the L_1 test for equality of variance described by Johnson (12), and a value of 0.978 was obtained. Entering Nayer's Table with $k = 3$ and $n = 60$, it is found that a value of L_1 equal to or less than 0.967 justifies rejection of the hypothesis of equal variance ($p = .05$). The three groups studied may thus be considered to have initially equal variance.

Product moment correlations between response magnitude and initial level were computed for each group at the first acquisition trial. Values of r were from .120 to $-.239$. None of these values is significant at the five per cent level. Thus, while there are considerable differences among the groups studied, response magnitude may be regarded as independent of the initial level from which it commenced. A final check on the approximate comparability of the anxiety, normal, and schizophrenic groups was made via analysis of variance of response magnitudes at the first acquisition trial. Since the value of F was but 0.14, it may confidently be concluded that all groups were responding initially with comparable magnitude.

Consequently, rates of extinction were studied via a two-way classification analysis of variance for repeated measurements. Table 2 shows mean response magnitude for each group by extinction trials. Each mean value was derived from the responses of the 60 *S*s in the group, each third of whom was partly extinguished to one of the three tones employed in the study. In other words, each mean value is actually a regression constant, *a*. It will be observed from Table 2 that there are fairly consistent trends showing the anxiety group to be responding with the greatest, the schizophrenic group with the smallest, and the normal group with intermediate magnitude. The analysis is shown in Table 3. The between groups (i.e., anxiety, normal, and schizophrenic) F ratio of 8.16 ($p < .001$) was obtained using the variance between *S*s in the same group as the error term. The between extinction trials F ratio of 235.21 ($p < .001$) and the Groups × Trials interaction F ratio of 7.98 ($p < .001$) were obtained using as error term the interaction for Pooled *S*s × Trials. These results indicate at highly acceptable levels of significance that by the eighth extinction trial all groups had shown a decrease in response magnitude. Further, the magnitude of response was consistently greatest in the anxiety group and smallest in the schizophrenic group, and the absolute intergroup differences in response magnitude were clearly an increasing function of the number of extinction trials. The data are thus in accordance with the original assertion that the anxiety group has the higher and the schizophrenic group the lower total drive level (Hull's *D*). They are in general accordance also with the relevant aspects of recent studies of anxiety and the extinction of conditioned responses (2, 22).

Since heterogeneity of variance among extinction data was not, however, precluded by application of Bartlett's test, thus leaving the acceptability of these results somewhat equivocal, a further test was performed with a nonparametric method devised by Wilson (28). This test has the advantage of being able to explore interaction as well as main effects without any supposition concerning the vari-

TABLE 2
GSR MAGNITUDES[a] DURING EXTINCTION TRIALS
(Log. conductance change + K)

Extinction Trial Number	Group		
	Anxiety	Normal	Schizophrenic
1	1.74	1.68	1.64
2	1.58	1.56	1.41
3	1.31	1.38	1.20
4	1.28	1.12	1.00
5	1.21	0.70	0.75
8	0.81	0.25	0.09

[a] Each value in the table is a rounded mean of 60 *S*s.

TABLE 3
ANALYSIS OF VARIANCE OF EXTINCTION DATA

Source of Variation	*df*	Mean Square	*F*	*p*
Between groups	2	8.770	8.16	<.001
Between *S*s in same group	177	1.074		
Between trials	5	39.752	235.21	<.001
Interaction: Trials X Groups	10	1.349	7.98	<.001
Interaction: Pooled *S*s X Trials	885	0.169		
Total	1079			

TABLE 4

CONTINGENCY TABLE FOR EXTINCTION DATA*

Trial	Group			
	Anxiety	Normal	Schizophrenic	$_bf_{i\cdot}$
1	7	9	7	23
2	13	14	22	49
3	30	20	25	75
4	30	39	32	101
5	32	54	46	132
8	41	59	60	160
$_bf_{\cdot j}$	153	195	192	$540 = n_b$

* Subscripts *a* and *b* indicate that the referent lies above or below the median, and $n_a = n_b$, and all n_{ji} are equal.

TABLE 5

SUMMARY OF NONPARAMETRIC ANALYSIS OF EXTINCTION DATA

Source	χ^2	*df*	*p*
Columns (Groups)	12.078	2	$<.01$
Rows (Trials)	289.960	5	$<.001$
Interaction G × T	29.018	10	$<.005$
Total	331.056	17	

ability of data. The median for the entire set of data was computed and a $2 \times r \times c$ contingency table (see Table 4) was set up, where $r = 6$ (Trials) and $c = 3$ (Groups), and the third factor corresponds to an exact division of the data into two sets, above and below the median. A total value of chi square was computed which was then broken down into main and interaction effects. The resultant analysis (Table 5) yielded a significant between groups value of 12.078 (with 2 *df*, $p < .01$), a significant between trials value of 289.96 (with 5 *df*, $p < .001$), and a significant interaction value of 29.018 (with 10 *df*, $p < .005$). While these results are *statistically* in general accordance with those obtained from the parametric method, there is one important consideration which leaves *interpretation* yet equivocal. Reference to Table 4 shows that when contingency data are used as an indirect measure of response strength during extinction, the normal and schizophrenic populations show virtually zero differences in relation to each other. The significant between groups and interaction values of chi square are thus probably due largely to the predicted deviation of the anxiety group from the other two. On the one hand, such an outcome is possible because of the additive properties of chi square; and on the other, it should be noted that considering the behavior measure employed in this study, resort to a nonparametric method potentially involves the loss of considerable information (cf. the relatively consistent smaller mean response magnitude in the schizophrenic group in Table 2).

Some further light was thrown upon the dubious normal-schizophrenic differences by a comparison of the 60 psychotic *S*s with 36 who were not conditionable because of zero GSR to shock, complete GSR adaptation to shock during the acquisition trials, or failure to give any response to the first and subsequent extinction trials. The unconditionable group had been given stronger shock ($p < .01$), showed a much higher resistance level ($p < .01$), were an average of 4.5 years older ($p = .05$), and had been ill for an average of seven years longer ($p < .01$). While certainly suggesting that the schizophrenic population was not randomly sampled, these facts enhance the significance of other results reported here.

DISCUSSION

Theoretical considerations led at the start of this work to a prediction of strongest conditioning in the anxiety group and weakest conditioning in the schizophrenic group. While the anxiety *S*s clearly showed significantly greater response magnitude during extinction compared with all other *S*s, it was not possible to demonstrate conclusively in any rigorous, formal fashion that the chronic schizophrenic *S*s showed significantly *smaller* response magnitudes than the normal *S*s. But there can be no doubt whatsoever that a significant degree of extinction occurred by the eighth extinction trial, that a significant interaction exists between groups and trials, and that the schizophrenic *S*s show a significantly lower response magnitude than the *anxiety* *S*s. These general conclusions are thus in accordance with recent studies of questionnaire-selected high and low anxious *S*s (e.g., 2, 22, 24).

At the same time, other recent predictions and conclusions (23, 25), such as the contention that the psychotic conditions autonomically more readily than both the normal *and* the neurotic, are contradicted by the results presented here. The prediction made for the behavior of a schizophrenic in a nociceptive conditioning situation will, however, more likely be a function of the degree of chronicity and perhaps the deteriorative character of the psychosis. Some of the psychotics studied by Spence and Taylor (23, 25) were outpatients; those here reported upon were all chronic, hospitalized cases, which may partly account for the obtained differences. But the comparison reported above of conditionable and unconditionable schizophrenic *S*s, showing a greater chronicity in the latter group, most strongly suggests a relationship between chronicity and autonomic conditionability.

In the light of the foregoing considerations, it is very difficult to make an overall statement about anxiety, psychosis, and Hull's *D*, unless the relative strengths of *D* can be roughly specified at different stages of the disorder. In other words, some psychotics *may* condition more readily than normals and even than anxious *S*s; but, a priori, the converse is just as likely, depending upon the various qualifications raised here.

SUMMARY

Conditioning of the GSR to electric shock in 60 hospitalized anxiety states, 60 normal controls, and 60 chronic functional schizophrenics yielded the following conclusions:

1. With magnitude of response (in log conductance units) during experimental extinction as an indirect measure of strength of conditioning, the anxiety *S*s showed significantly stronger conditioning than both the normal and schizophrenic *S*s, the difference in response magnitude being some increasing function of the number of extinction trials, as shown by a significant interaction term. While schizophrenic *S*s showed the lowest *magnitude* of response, it was not possible to offer proof with either a parametric or a nonparametric method of analysis that this magnitude was statistically less than that for the normal *S*s.

2. It was argued that predictions concerning the autonomic conditionability of the schizophrenic must vary with the degree of chronicity of and perhaps the amount of deterioration in the disorder, with less reactivity (for example, Hull's drive strength, *D*) being more typical of the chronic and deteriorated case. A comparison of schizophrenic *S*s who were conditionable with some who were not gave support to such an hypothesis.

REFERENCES

1. BENDER, LAURETTA, & SCHILDER, P. Unconditioned and conditioned reactions to pain in schizophrenia. *Amer. J. Psychiat.*, 1930, **87**, 365–384.
2. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
3. HAGGARD, E. A. On the application of analysis of variance to GSR data: III. Some effects of the use of inappropriate measures. *J. exp. Psychol.*, 1949, **39**, 861–867.
4. HILGARD, E. R., JONES, L. V., & KAPLAN, S. F. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–100.
5. HILL, D., ROE, P., THEOBALD, J., & WADDELL, M. A central homeostatic mechanism in schizophrenia. *J. ment. Sci.*, 1951, **87**, 111–131.
6. HOAGLAND, H. Adventures in biological engineering. *Sci.*, 1944, **100**, 63–67.
7. HOAGLAND, H. Enzyme kinetics and the dynamics of behavior. *J. comp. physiol. Psychol.*, 1947, **40**, 107–128.
8. HOAGLAND, H., & PINCUS, S. Steroid excretion and stress of flying. *J. aviat. Med.*, 1943, **14**, 173–193.
9. HOSKINS, R. G. *The biology of schizophrenia.* London: Chapman, 1946.
10. HOVLAND, C. I. Generalization of conditioned responses: II. Sensory generalization of CRs with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
11. HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
12. JOHNSON, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.
13. LACEY, O. L., & SIEGEL, P. S. An improved potentiometric circuit for measuring the GSR. *Amer. J. Psychol.*, 1948, **61**, 272–274.
14. MALMO, R. B., SHAGASS, C., & DAVIS, J. F. Electromyographic studies of muscular tension in psychiatric patients under stress. *J. clin. exp. Psychopath.*, 1951, **12**, 45–66.
15. MAYS, L. L. Studies of catatonia: V. Perseverational tendencies in catatonic patients. *Psychiat. Quart.*, 1934, **8**, 728–735.
16. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
17. MILLER, N. E. Comments on theoretical models illustrated by the development of a theory of conflict. *J. Pers.*, 1951, **20**, 82–100.
18. PAINTAL, A. S. A comparison of the GSR in normals and psychotics. *J. exp. Psychol.*, 1951, **41**, 425–428.
19. PFAFFMAN, C., & SCHLOSBERG, H. The conditioned knee jerk in psychotic and normal individuals. *J. Psychol.*, 1936, **1**, 201–205.
20. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
21. SHIPLEY, W. C. Studies of catatonia. VI. Further

investigation of the perseverational tendency. *Psychiat. Quart.*, 1934, **8**, 736–744.

22. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
23. SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.
24. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
25. TAYLOR, JANET A., & SPENCE, K. W. Conditioning level in the behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497–503.
26. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83–91.
27. WELCH, L., & KUBIS, J. Conditioned PGR (Psychogalvanic Response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372–381.
28. WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

Received October 31, 1956.

PARENTAL EVALUATIONS OF CHILDREN AND CHILDREN'S SELF-EVALUATIONS

MALCOLM M. HELPER

Nebraska Psychiatric Institute, University of Nebraska College of Medicine

SELF-EVALUATIVE attitudes are widely believed to arise rather directly from evaluations made of the individual by others, with especial influence frequently being attributed to evaluations made by parents (1, 9). In spite of this apparent theoretical agreement, or perhaps because of it, very little empirical study has been made of the relationships between parental evaluations of the child and the child's self-evaluations. Jourard and Remy (6) have made what appears to be the only study having direct bearing on this question. They studied relationships between college students' satisfaction with self and body (self- and body-cathexis) and parental evaluations of the students, the latter being reported by the students themselves. The correlations ranged from .56 to .70 between self-satisfaction expressed by the students and parental satisfaction with the students as reported by the students. These authors note, however, that parental evaluations should be assessed directly before definite conclusions are drawn, because reports by the students as to their parents' attitudes may be subject to systematic distortion.

In the present study, independent assessment was made of self-evaluation in a group of 8th and 9th grade children and of evaluation of the children by their parents.

METHOD

Forty-two adjectival rating items were developed on the basis of dimensions of personality description found by Cattell to be representative of a very large number of trait names (2). Four scales were also selected from Osgood's analysis of connotative meaning (7). It was believed that these bases for selection would insure a wide sampling of behavior descriptions and avoid undue emphasis on any one aspect of personality description. Each scale was defined at the two extremes by adjectives or adjectival phrases and offered a seven-point scale for rating, as in the example below:

talkative:__:__:__:__:__:__:__:quiet.

Subjects were asked to rate several concepts on each of the forty-six scales by checking the space that best described that concept as they viewed it. Each child was asked to rate himself as he ordinarily thought of himself (actual self-concept), and to rate himself as he would most like to be (ideal self-concept). Each parent used the scale to rate his child as he ordinarily thought of the child (actual child-concept) and to rate his child as he (the parent) most wanted the child to be (ideal child-concept). Other ratings were collected at the same time, and were utilized in investigating other aspects of children's self-description (5).

Subjects

Data were gathered on the 74 children in the 8th and 9th grades in the secondary school operated by the University of Illinois. One or both of the parents of 53 of the children cooperated in providing the ratings described above. Because of incomplete ratings on the part of a few *S*s, the number of families included in different computations varies from 50 to 53. Each computation included all of the families providing pertinent data. Since over one-half of the fathers involved were employed in an academic capacity by a university, the families must be considered to be primarily representative of a narrow portion of the socioeconomic spectrum.

Measures of Evaluative Attitudes

Two methods of measuring attitudes toward (or "acceptance" of) the self and others find current usage. In one method, items are selected that unequivocally indicate a favorable or unfavorable attitude toward the subject in question (3, 8). Ratings on these items can be summed into a single score since the favorability indicated by a given response has been determined in advance. This method has the advantage of simplicity of scoring but is limited to items that have unambiguous evaluational implications. In the second and more popular method, "actual" ratings on a group of items are subtracted from "ideal" ratings on the same group of items. The differences are summed across items to yield a score that reflects the extent of the discrepancy between the actual and ideal concepts. This method (of which the Q sort is a variant) allows for differences among *S*s in evaluating the desirability of a given trait. However, the difference scores are necessarily less reliable than the raw ratings from which they are derived (4).

In the present study, both types of measures were applied to children's self-ratings and to the parents' ratings of their children. A Favorability score was computed by summing "actual" ratings on 15 items on which there was very high agreement among raters as to desirability. These were items on which the ideal self-ratings were heavily concentrated toward one end of the rating scale. All items having a mean ideal self-rating more than 1.5 scale points from the midpoint of the 7-point scale were identified. Of these, the fifteen items having the smallest standard deviation in ideal self-ratings were selected for the Favorability scale. In practical terms, the distribution of ideal ratings on these items is so narrow that subtracting the actual ratings from them would add little information. The Favorability score was simply the sum of the raw ratings on these fifteen items after all items had been aligned so that a high rating indicated a desirable standing. A Child-Favorability score was obtained from each parent's actual child ratings, while a Self-Favorability score was obtained from each child's actual self-ratings.

A Self-Acceptance score was derived for each child from the actual-ideal discrepancies generated by the remaining 31 items. For each child, the 31 discrepancy scores were squared and then summed in accordance with the specifications for D, the statistic of profile comparison (4). The discrepancies on these items between each parent's actual-child and ideal-child ratings were treated in like fashion to yield a Child-Acceptance score. In this method of scoring, the higher the score, the less the Acceptance. Neither the Self-Acceptance nor the Child-Acceptance scores were distributed normally. For this reason, nonparametric statistical procedures were used in all analyses.

An estimate of each child's acceptability to his peers was available from a sociometric study made by the two classroom teachers involved. Each child had been asked to list classmates with whom he would like to work on a study project. The number of choices received by each child was taken as the index.

The data described above permit study of several questions relating to the problem of self-evaluation. For methodological purposes, the degree of correspondence between the two approaches to the assessment of self-evaluation was studied. The question of primary interest—the degree of correlation between parental evaluations of children and children's self evaluations—was investigated in terms of both the Favorability and the Acceptance variables. The relationship between children's self-evaluations and acceptability to peers was also considered.

RESULTS

Reliability coefficients were computed by the split-half method for Favorability and for Acceptance in both parents and children. Rank-difference correlation was used because of the non-normality of the Acceptance scores. These coefficients, after correction for total length by the Spearman-Brown formula, are presented in Table 1. The reliabilities are seen to be only moderate, but do appear high enough to justify further study of the measures.

The correlations between the two measures of attitudes toward the child, shown in Table 2, are much lower than their respective reliabilities. The two measures do not appear to tap the same aspects of evaluative attitudes toward the child, either in the parents or within the child himself.

TABLE 1

RELIABILITY OF FAVORABILITY AND ACCEPTANCE SCORES[a]

Group	N	Favorability	Acceptance
Boys	34	.86	.77
Girls	40	.77	.80
Fathers	53	.66	.80
Mothers	51	.85	.83

[a] Coefficients are rhos corrected by the Spearman-Brown formula for total length of the scales.

TABLE 2

INTERCORRELATIONS OF FAVORABILITY AND ACCEPTANCE MEASURES

Group	N	Rho
Boys	34	.24
Girls	40	.21
Fathers		
Of Boys	20	.10
Of Girls	30	.18
All Fathers	50	.16
Mothers		
Of Boys	21	.42
Of Girls	30	.23
All Mothers	51	.39*

* Significant at the .01 level.

Rank difference correlations between parental Favorability and Acceptance scores and the corresponding self-evaluative scores in children are presented in Table 3. In general, the parent-child correlations are low for both the Favorability and the Acceptance variables. Of the 24 correlations computed, 5 reach significance at the .05 level, 4 fall between the .05 and .10 levels, and 15 do not approach statistical significance. All except one of the correlations fell in the predicted positive direction. (It is not claimed, of course, that the 24 correlations are entirely independent.) On the whole, then, the data do seem to point to the existence of a slight but real tendency toward similarity between parents' evaluations of their children and the children's self-evaluations.

While the correlations do not differ enough among themselves to warrant differential interpretations of the various parent-child combinations, it can be noted that all but one of the significant and near-significant correlations are between corresponding variables at the parent and child levels. That is, parental Favorability has significant relationships only with Self-Favorability in the child, while parental Acceptance is related much more

TABLE 3
PARENT-CHILD CORRELATIONS (RHOS) FOR TWO MEASURES OF EVALUATION OF THE CHILD

	Boys			Girls			All *S*s		
	N	Favor-ability	Accept-ance	*N*	Favor-ability	Accept-ance	*N*	Favor-ability	Accept-ance
Fathers									
Favorability	20	.44*	.26	30	.26	.06	50	.32**	.11
Acceptance	20	.18	.42*	30	.28	.44**	50	.31**	.37**
Mothers									
Favorability	21	.20	.10	30	.33	.18	51	.22	.08
Acceptance	21	.29	.08	30	−.15	.39**	51	.16	.27*

* Coefficient reaches .10 level, two-tailed test.

** Coefficient reaches .05 level, two-tailed test.

consistently to children's Self-Acceptance than to children's Self-Favorability.

The small size of the parent-child correlations does not rule out the possibility of more substantial relationships in certain subgroups. In view of the low correlations between parental Acceptance and Favorability, it would seem reasonable to study the possibility that different combinations of the two variables may be associated with differences in children's self-evaluations. One could reason that children whose parents view them unfavorably would be rather highly sensitive to their parents' acceptance of them and their faults, even though the children's Self-Favorability does not reflect the parents' basically negative view of them to any great extent. On the other hand, children whose parents view them as basically sound might well be less sensitive to any unusually high aspirations (high actual-ideal discrepancy) that the parents might express for them. In terms of the present variables, there should be a relatively high relationship between parental Acceptance and children's Self-Acceptance (or Self-Favorability) in cases where parents have low Favorability scores, but not in cases where parents have high Favorability scores.

In accordance with this hypothesis, the relationship between parental Child-Acceptance and the two self-evaluative variables was studied separately in the group of families characterized by high parental Favorability and in the group characterized by low Favorability. The median Favorability score was used as the dividing line. Separate analyses were carried out for mothers and for fathers. In each low Favorability group, children receiving high and low parental acceptance were compared with respect to Self-Acceptance and Self-Favorability. Similar comparisons were made in each high paternal Favorability group. The Mann-Whitney *U* test was used in making all statistical tests. The results of this procedure can be described briefly: The one significant difference occurred in the group characterized by low Favorability of the mother. Here, as expected from the hypothesis, children receiving low Acceptance from the mother had lower Self-Acceptance than did those receiving high Acceptance from the mother. However, the preponderance of nonsignificant differences indicates that, in terms of the present data, the hypothesized effects are slight at best, and probably should be discounted completely without further evidence.

TABLE 4
COMPARISON OF HIGH AND LOW SOCIOMETRIC GROUPS ON SELF-EVALUATIVE VARIABLES
(U-test)

			Self-favorability		Self-acceptance	
	N	*N*[a]	Sum of ranks	*p*	Sum of ranks	*p*
Boys	34	15	221.5	—	160.5	.001
Girls	40	22	423.0	—	466.0	—

[a] Number of cases in low sociometric group.

To study the relationship between children's self-evaluations and their acceptability to their peers, the distribution of choices received in the sociometric procedure was first dichotomized for each sex separately. For each sex, the *U* test was employed to evaluate the difference in Self-Favorability and Self-Acceptance between the more and the less popular children. These results are presented in Table 4. The popular boys express a significantly higher degree of Self-Acceptance than the unpopular boys. No significant differences emerge for girls.

DISCUSSION

In interpreting the rather tenuous relationship observed between parents' evaluations of children and the children's self-evaluation, we should first take into account the problem of the validity of the measures used. While the children's verbal responses can, if desired, be taken as operational definitions of their self-evaluative traits, the parental variables specified by the theoretical position mentioned at the outset are those which reflect actual evaluative behavior by the parent toward the child over a span of several years. From this point of view, it is rather encouraging that an apparently significant relationship does emerge when only rating data are available from the parents.

The contrast between the present low correlations and the relatively high correlations reported by Jourard and Remy (6) can be interpreted in two basically different ways. It is possible, as those authors note, that the children's reports of their parents' attitudes are distorted (possibly by response set) in such a way as to coincide spuriously with the children's self-ratings. On the other hand, it is also possible that children's reports are actually more valid as measures of parental evaluations than ratings by the parents themselves. This uncertainty would seem to call for validation of both types of measures against observed parental evaluative behavior.

Regardless of the size of the true correlations between parental attitudes and children's self-evaluations, no causal interpretations are justified until temporal trends in the two variables can be studied. This type of study would represent a logical step from the presently available evidence of a positive relationship between parental evaluations of the child and the child's self-evaluations. Although psychodynamic interpretations are premature at the present time, it does seem wise, in view of the small size of the correlations found, to note that parental evaluations may have very great consequences for the child's personality even if they are not reflected directly in the child's self-evaluations. In fact, a study of the reactions of children whose parents evaluate them negatively might well shed much light on defensive processes.

The evidence that Acceptance and Favorability are substantially independent variables is rather difficult to interpret theoretically since the two types of measures were originally distinguished on methodological grounds and were expected to reflect the same attitudinal variables. It was suggested in comparing the effects of the various combinations of parental Favorability and Acceptance that Favorability reflects simply a description of an individual in evaluative terms while Acceptance indicates the attitude toward the individual with his perceived defects or assets. This differentiation would suggest that Acceptance is the more emotionally loaded of the two variables and has more significance for interpersonal relationships. While the data of the present study do not resolve this question, it can be noted that the single significant relationship between self-description and sociometric status involved the Acceptance variable. While the present results provide only the scantiest basis for interpretation of these two aspects of self-evaluations, they do point up the need for further methodological study of such constructs as self-acceptance and self-esteem.

SUMMARY

Two self-evaluative measures, labelled Self-Favorability and Self-Acceptance, were derived from self-ratings by 74 8th- and 9th-grade children. Analogous measures of parental Favorability and Acceptance with respect to the child were derived from ratings by parents of 51 of the children. Sociometric information was available for the children. Although the measures of Favorability and Acceptance had adequate reliability, correlations between the two variables were low in both parents and children. Correlations between parental evaluations and children's self-evaluations tended to be small but consistently positive. The results were interpreted as evidence that attempts to increase the validity of assessments of parental attitudes and further to differentiate aspects of self-evaluations would prove rewarding.

REFERENCES

1. CAMERON, N., & MAGARET, A. *Behavior pathology.* New York: Houghton Mifflin, 1951.
2. CATTELL, R. B. *Personality.* New York: McGraw-Hill, 1950.

3. CRANDALL, V. J., & BELLUGI, U. Some relationships of interpersonal and intrapersonal conceptualizations to personal social adjustments. *J. Pers.*, 1954, **23**, 224–232.
4. CRONBACH, L. J., & GLESER, G. C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
5. HELPER, M. M. Learning theory and the self concept. *J. abnorm. soc. Psychol.*, 1955, **51**, 184–194.
6. JOURARD, S. M., & REMY, R. M. Perceived parental attitudes, the self, and security. *J. consult. Psychol.*, 1955, **19**, 364–366.
7. OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.
8. SOLLEY, C. M., & STAGNER, R. Effects of magnitude of temporal barriers, type of goals and perception of self. *J. exp. Psychol.*, 1956, **51**, 61–70.
9. SNYGG, D., & COMBS, A. W. *Individual behavior.* New York: Harper, 1949.

Received December 19, 1956.

IDENTIFICATION IN GROUP THERAPY[1, 2]

GERALD J. BRISKIN

Byron Harless and Associates, Miami, Florida

THE process of psychotherapy implies changes in behavior, and the variables associated with these changes have become a focal point for research in therapy. In a recent study, Schrier (11) found that identification with the therapist in individual therapy was significantly related to therapeutic success. This suggests that the ability to identify with the therapist is an important variable in the prognostic evaluation of candidates for therapy. The intent of this study is to investigate two of the variables contributing to identification.

For the purposes of this paper, identification is operationally defined as changes in patient behavior from pretherapy status to behavior similar to that manifested by the therapist. The variables selected for study were rigidity and lability.

Rigidity is a well-studied entity. Considerable research indicates that certain individuals have difficulty in changing their patterns of behavior. They seem encysted against the world and behave in a stereotyped and inflexible manner. Tension and anxiety are well "bound" in this personality configuration, and emotional content is effectively removed from thoughts and ideas to the extent that anxiety is not strongly experienced by this kind of individual in therapy. Fenichel (5) suggests that identifications are always motivated by drives or tension. Since the rigid individual may not experience tension very readily, and since access to and from the ego is limited, it is quite likely that he will have difficulty in identifying.

The term lability, used in this context, refers to a personality pattern that closely approximates the preschizophrenic personality in which the ego boundaries are vaguely defined and weakly maintained. An individual possessing this kind of ego structure has difficulty in distinguishing between self and non-self. Because of this, his reality-testing abilities are limited, and he often confuses events occurring in the external world with events occurring in himself. As a result, his ability to identify is severely impaired.

Based upon this reasoning, two hypotheses were formulated: (*a*) There is an inverse relationship between rigidity and identification; (*b*) there is an inverse relationship between lability and identification.

METHOD

The *S*s utilized in this study were ninth- and tenth-grade male and female high school students. They were selected for the experiment by asking members of the social living classes if they would like to volunteer to participate in groups that would discuss personal problems of adolescents. They were told that only a limited number could be selected and that they would be excused from class to participate. A total of 109 students volunteered. All of the volunteers were then rated by their social living class teachers on a form designed to assess emotional adjustment. Those students who were rated most poorly adjusted were then eliminated from the study on the ground that they would be least likely to benefit from short term therapy.

Three groups were selected to participate in therapy, each composed of five boys and five girls. Before the therapeutic groups were formed, *S*s were asked to fill out a specially designed questionnaire,[3] and were given group forms of the Rorschach. These forms and tests were administered by members of the research staff other than the therapists.

Following this collection of data, the groups started in therapy, each group having a different therapist. The groups met twice a week over a period of four months, with interruptions for school holidays. *S*s were told that they were free to drop out of the groups up to the fifth session, but they were expected to stay in the group after the fifth meeting. There was no other pressure to compel attendance.

The therapeutic approach followed group analytic principles, with the therapist very active in serving as the focal point of the group, yet being essentially per-

---

[1] Based upon a Ph.D. dissertation written at the University of Michigan. The author wishes to express his appreciation to Max L. Hutt, Donald Marquis, Theodore Newcomb, Edward Bordin, and Alvin Zander, who contributed much to the thesis. Gratitude is also due the Rackham Fund, whose grant supported the Group Therapy Project on which this work is based.

[2] Paper delivered at the APA, September, 1954.

[3] To save printing costs, a 3-page tabular presentation of the questionnaire has been deposited with the American Documentation Institute. Order Document No. 5465 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

missive and accepting of all behavior, including fairly extreme forms of acting out. This technique is based on Wender's (13) notions of group therapy and is designed to elucidate and clarify defensive behavior and its significance.

At the conclusion of therapy, all *S*s again filled out the same questionnaire that they had at the beginning of the study. These questionnaires were administered by the therapists.

Each of the therapists was rated on this same questionnaire by experienced third- or fourth-year clinical psychology graduate students who knew them well but who were not involved in any competitive relationships with them. These judges were instructed to answer the questionnaire as they believed the therapist would. A combined rating for each therapist was then obtained by averaging the judges' ratings.

Having the therapists rated by judges seemed a sounder procedure than having the therapist rate himself or having him rated by his group members. With regard to the first alternative, the therapist, regardless of his insight, may still tend to rate himself defensively and thus present a distorted picture of himself. Cutler (4) made use of this reasoning in utilizing discrepancies between therapists' self-ratings and observer ratings as an index of therapists' areas of conflict. The other alternative, on casual examination, seems to have considerable merit. However, since *S*s in therapy may be involved in all sorts of transference relationships with the therapist, any ratings of him by them may also be distorted. The use of external observers was therefore decided upon.

The questionnaire defining identification in therapy derived from Murray's (8) inventory was altered to fit the needs of an adolescent group. It consisted of 50 statements, each of which could be answered *always*, *at times*, *seldom*, *never*. Each item was a description of behavior or feelings—for example: "I feel I should like to be a leader of a group," or "I try to be good in home and in school."

The Rorschach test was essentially a group adaptation of the individual administration. The Harrower-Erickson Rorschach slides were used. The protocols were scored for Beck's (1) locations, content, form levels, and populars and for Klopfer's (7) determinants. Since group Rorschachs differ considerably from individual Rorschachs, especially as to the inquiry, it was necessary to adhere to certain additional rules: If a percept was named in the free association but there was no inquiry or location, it was not scored. If a percept was named in free association and there was a location but no inquiry, it was scored F only for determinants. Precision alternates were scored only once. Otherwise, the records were scored in accordance with the usual rules specified by Klopfer and Beck. No reliability study of the scoring was done since numerous reports in the literature (10, 12) indicate that high interjudge reliability can be expected between Rorschach scorers of considerable experience.

The measure of identification utilized in this study was operationally defined as the total number of shifts from each *S*'s pretherapy questionnaire to his post-therapy questionnaire in the direction of the therapist as rated by the observers. The measures of lability and rigidity were derived from the Rorschachs. However, since the determinants obtained from a group administration of the Rorschach vary considerably from those obtained in individual administration, it was not possible to apply standard interpretive procedures to the Rorschach records without first evaluating the data. In order to accomplish this evaluation, the Rorschachs for the volunteers who did not receive therapy were scored and distributions made of all determinants. Since most of these distributions were greatly skewed, medians and quartile deviations were calculated rather than means and standard deviations. These statistics were then used as standards in setting up the rating scales.

However, Rorschach variables do not scale proportionately; a total of four M responses, for example, is not twice as significant as two M responses. Therefore, in assigning weights to Rorschach variables, cutting points must be determined in accordance with the shape of the distribution of the variables dealt with. Furthermore, since most Rorschach work has been done with individually administered tests, the "accepted" cutting points must also be modified in terms of the group administered distributions. An example follows:

It is a standard Rorschach procedure to regard a protocol in which F comprises more than fifty per cent of the responses as constricted. But the median F percentage for the group records was 74.5. Two explanations are available: The population dealt with produced unusually frequent F percepts, or the lack of an individually conducted inquiry resulted in many responses being scored F which otherwise would not have been so scored. Since the latter explanation seems more tenable, the notion of the cutting point for constriction as 50 per cent was discarded, and the cutting point used was based on Q3, which was 88. Thus, a weighted scale for constriction could be established with scores of 87 per cent or below weighted at zero, 88–95 per cent at one, and 96 per cent or over at two.

Based on this procedure, the rating scales for lability and rigidity were developed. To determine the rigidity or lability scores for a particular *S*, the determinants from his Rorschach were given the appropriate factor weights as indicated in the respective scales, and these weights were then totaled. Since some negative scores were obtained, all of the scores had a constant added to them so that positive numbers could be dealt with.

The measure of rigidity was based upon pertinent theoretical and clinical positions as found in the literature (2, 6, 7, 10). One study of particular relevance by Cowen and Thompson (3) related certain Rorschach findings to measures of Einstellung rigidity. The rigidity index may be found in Table 1.

As discussed previously, the measure of lability is closely related to what has been described as the preschizophrenic personality. Consequently, the scale measuring this configuration was derived from standard Rorschach "signs" (2, 6, 7) considered as indicative of such a personality. These signs and their weights are given in Table 2.

The interobserver agreement of ratings of the therapist can be found in Table 3. These agreements are given in terms of coefficients of contingency rather than the usual product-moment correlations. This statistical treatment was necessitated by the nature of

TABLE 1

The Rigidity Rating Scale

Factor	Weight
I. Total Responses	
11 or less	3
27 or more	3
II. D plus Dd%	
85 or more	3
25 or less	3
III. Sum M	
0	3
1	0
2 or more	−3
IV. F%	
85 or more	3
50 or less	−3
V. F + %	
90 or more	3
60 or less	−3
VI. A%	
60 or more	3
30 or less	−3
VII. Number of Content Categories	
4 or less	3
9 or more	−3
VIII. Populars	
6 or more	3

TABLE 2

The Lability Rating Scale

Factor	Weight
I. Sum CF plus C greater than FC (FC = .5, CF = 1.0, C = 1.5)	
by 0–.5	0
by 1.0–2.0	2
by more than 2.0	4
II. Total M	
0	3
1	0
2 or more	−3
III. F% plus F + %	
F% greater than 80, F + % less than 65	6
IV. Total FM	
0–1	0
2 or more	3
V. M quality	
For each flexor	−2
For each extensor	2
VI. Pure H% (does not include human like percepts)	
15 or less	3
30 or more	−3
VII. Card Rejections	
0–1	0
2 or more	3

TABLE 3

Coefficients of Contingency*** for Interjudge Agreement in Rating Therapists

Therapist A	Therapist B	Therapist C
.36**	.40**	.36*

* .05 level of confidence.
** .01 level of confidence.
*** Upper limit of coefficient of contigency = .7071.

the data. Ratings on the questionnaire could be made only in terms of four categories, *always*, *at times*, *seldom*, *never*. Since the distributions of these ratings were skewed, and since product-moment correlations are attenuated by a limited range (only four categories in this sample), the usual correlational techniques were inappropriate and, consequently, the contingency coefficient was substituted. A fourfold table was used, grouping *always* and *at times* in one category and *sometimes* and *never* in the other.

Results

The hypothesis specifying the relationship between identification and rigidity was tested by correlating the identification scores of *S*s in the therapy groups with their rigidity scores. A product-moment correlation of −.73 was obtained, significant at the .01 level. In order to test the possibility that rigid individuals might show less identification with the therapist simply by making fewer shifts on the questionnaire, regardless of direction, as nonrigid *S*s, the total number of shifts for each *S* was correlated with his respective rigidity score. The product-moment correlation was .09. A *t* test between the differences of the mean number of shifts for the rigid and nonrigid group was similarly not significant. The possible objection therefore seems irrelevant, and the findings may be interpreted as supporting the hypothesis.

The hypothesis specifying the relationship between identification with the therapist and lability was also tested by correlating the identification score of each individual in therapy with his lability score. A product-moment correlation of −.64 was obtained, significant at the .01 level. This finding is also interpreted as supporting the hypothesis.

Discussion

The primary intent of this study was to make an exploratory attempt to evaluate two hypotheses derived from psychoanalytic theory as they apply to group therapy. In general, both hypotheses were supported. These findings lend emphasis to the role that ego structure plays in identification and to the fruitfulness of research on ego functions.

Previous studies have suggested that the ability of patients to identify with the therapist markedly increases the probability of therapeutic success. Since there are far more prospective patients than the limited number of psychotherapists can accommodate, the selection of candidates for therapy is generally made on the basis of an estimated ability to benefit from treatment. In this respect, the present findings imply potentially useful criteria for the prediction of identification, which may be a highly relevant factor in prognostic evaluation.

SUMMARY

The *S*s of this study were male and female adolescents (12–15 years of age) who were asked to volunteer to participate in groups that "would discuss adolescent problems." Those who volunteered were rated by their teachers on emotional adjustment. *S*s rated as most poorly adjusted were eliminated from the study. Three therapy groups were organized, each composed of five boys and five girls, each in the charge of a different therapist. The members of the group were given a special questionnaire, designed to assess needs, and a group form of the Rorschach before therapy began. At the conclusion of therapy, all *S*s again took the questionnaire. Two or more judges rated each therapist on the questionnaire items.

A measure of identification was obtained by taking the number of shifts by the *S*s from pre- to posttherapy questionnaires which were in the direction of the therapist as defined by the combined observer ratings. Measures of lability and rigidity were derived from the Rorschach tests.

The hypotheses that both rigidity and lability were inversely related to identification were tested by correlational methods. Results were positive in both instances.

REFERENCES

1. BECK, S. *Rorschach's test: Basic processes.* Vol. I. New York: Grune & Stratton, 1949.
2. BECK, S. *Rorschach's test: Advances in interpretation.* Vol. III. New York: Grune & Stratton, 1952.
3. COWEN, E. L., & THOMPSON, G. G. Problem solving rigidity and personality structure. *J. abnorm. soc. Psychol.*, 1951, **46**, 165–176.
4. CUTLER, R. The relationship between therapists' personality and certain aspects of psychotherapy. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
5. FENICHEL, O. *The collected papers of Otto Fenichel.* First Series. New York: W. W. Norton, 1953.
6. KLOPFER, B., AINSWORTH, M. D., KLOPFER, W. G., & HOLT, R. R. *Developments in the Rorschach technique.* Vol. I. New York: World Book, 1954.
7. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique.* New York: World Book, 1946.
8. MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
9. RORSCHACH, H. *Psychodiagnostics.* Berne: Verlag Hans Huber, 1949.
10. SCHNEIDER, S. F. Prediction of psychotherapeutic relationship from Rorschach's test. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
11. SCHRIER, H. The significance of identification in short-term therapy. Unpublished doctoral dissertation, Univer. of Michigan, 1951.
12. SILVERMAN, H. The prediction of consciousness of conflict in the self from the Rorschach. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
13. WENDER, L. The dynamics of group psychotherapy and its applications. *J. nerv. ment. Dis.*, 1936, **84**, 54–55.

Received February 1, 1957.

SOME RELATIONSHIPS AMONG PROJECTIVE TEST, INTERVIEW, AND SOCIOMETRIC MEASURES OF DEPENDENT BEHAVIOR[1]

BERNARD J. FITZGERALD[2]

VA Hospital, Roanoke, Virginia

THE need for experimental evaluation of projective test data as used for describing and evaluating behavior has frequently been pointed out (4, 7). Some authors disparage such experimental evaluation, suggesting that perhaps clinically useful evidence is distorted or lost in the process of quantification (2, 3). However, it would seem that systematic investigation of projective test data is essential, particularly in view of the failure of many studies to predict successfully from projective test responses to overt behavior, even on the basis of clinical evaluation of the data.

The present study was developed using descriptive constructs from Rotter's social learning theory (6). From the point of view of social learning theory, any given situation has within it possibilities for the satisfaction of many different needs. Which of these needs an individual will strive to satisfy will depend upon two factors, the *value* that the satisfaction of a given need has for him and his *expectancy* that his behavior will lead to that satisfaction. The value which the satisfaction of a given need has for the individual is termed Need Value. In the measurement of Need Value, emphasis is upon what need satisfactions the individual *prefers*. An individual's expectations that particular behaviors will lead to the desired need satisfaction is termed Freedom of Movement. Emphasis is upon what the individual *expects* to happen. Freedom of Movement is high if the individual expects his behaviors to lead to need satisfaction. An individual's Freedom of Movement is low if he expects failure or punishment for such behaviors. One method of measuring Freedom of Movement depends upon identifying the frequency with which the individual, when seeking any particular need satisfaction, resorts to symbolic rather than direct, overt behavior to obtain that satisfaction; the more he resorts to symbolic behavior, the lower is his Freedom of Movement.

In terms of this study, it was hypothesized that if projective test responses represent, in part, symbolic ways of obtaining gratification, then frequency of response should be highest when Freedom of Movement is low and Need Value high. This hypothesis assumes that a projective test provides a situation in which symbolic gratification is possible; therefore, the individual tends to give responses involving satisfactions that have a high value for him but for which he has a low expectancy of realization. Such a situation, where Need Value is higher than Freedom of Movement, is considered in this study to be a situation of important conflict.

The purpose of this study was to provide data related to the interpretation of needs as reflected in the content of projective test responses. Dependency, the need investigated, may be defined as the need to have other individuals prevent frustration or punishment and provide for the satisfaction of other needs. The projective tests used were the Murray TAT and the Rotter Incomplete Sentences Blank (ISB). The measures of overt behavior were a sociometric technique and an interview. The measures of Need Value and Freedom of Movement for Dependency were also derived from the interview.

The hypotheses of this study are as follows:

1. Frequency of n Dependency responses in a projective test is positively related to sociometric and interview indices of Dependent behavior; and
2. Frequency of n Dependency responses in a projective test is positively related to the importance of conflict in n Dependency.

METHOD

Eighty *Ss*, 20 from each of four social fraternities, participated in the study. All *Ss* had been members of their respective fraternities for at least six months and students at the Ohio State University for at least one year. A pilot study was devoted to the development of the interview, the sociometric measure, and judges' manuals for scoring interview and projective test responses.

---

[1] This paper is based on a doctoral dissertation done at the Ohio State University. The author acknowledges with gratitude the advice and guidance given him by J. B. Rotter, under whose direction this study was carried out. The writer also is grateful to Gordon Rader for his helpful criticism of this manuscript.

[2] Now at Philadelphia State Hospital.

The criteria of overt behavior were a sociometric index and an interview measure of Dependent behavior. A nominating technique was used as the sociometric measure. Two descriptions were provided the *S*s in order to make their nominations. One of these descriptions was worded to describe certain behaviors considered most and another worded to describe behaviors least representative of Dependent behavior. Care was taken not to label the descriptions as representative of any particular kind of behavior, e.g., Dependent. *S*s were asked to select from the members of the group those who, in general, best fit one or the other of the two descriptions. The nominations were made alternately, i.e., the one *S* who behaved most Dependently, the one *S* who behaved most non-Dependently, and so on. In this manner, then, all *S*s made three nominations for each description.[3]

A structured clinical interview was developed to elicit from these same *S*s historical data relevant to their Dependent behavior and data relevant to their Need Value and Freedom of Movement for Dependency satisfaction. The interview was divided into eleven areas: school, home, advisor, major academic subject, roommate, friends, girl friends, health, personal problems, and religion. A manual for scoring the interview variables of Dependent behavior, Need Value, and Freedom of Movement for Dependency satisfaction was developed. Included in the manual were the definition of Dependency and a set of general principles derived from the definition, as well as examples from interview protocols obtained in the pilot study. Examples were derived from these interviews to correspond to the points 1, 4, 7, and 10 on a ten-point rating scale. The in-between points were included to allow finer discriminations than were possible from the selected referents in the manuals. The values attached to the examples in each of the manuals were arrived at by the experimenter's first assigning crude values consistent with the context from which the examples were taken. Two independent raters also assigned values to these examples, and the scale values were adjusted to maximize agreement among the three raters. Finally, these refined scales were discussed with the interview judges. Changes were made in the scales where necessary for complete agreement among the judges, and the examples were then assigned final values.

Judges' scoring manuals were similarly developed for the TAT (Cards 3BM, 4, 6BM, 7BM, 8BM, 10, 12M, 13MF, 15, 17BM, 18BM, 20) and the ISB. From the TAT stories obtained in the pilot study, a manual was prepared to aid in the determination of the presence or absence of a Dependency plot for any given story. In the development of the manual, each pilot protocol was carefully read, and any story which could be considered to have a Dependency plot was noted. Two projective test judges and the experimenter discussed those stories which were "borderline" or difficult to judge for n Dependency. Following discussion of each borderline story, the three judges agreed on each story as representative of either Dependency or non-Dependency. These stories, as well as stories clearly representative of n Dependency, were included in the final form of the manual. Dependency plots were identified in relation to both the Hero and the Other Figures. Suggestions were included in the manual to promote agreement among the judges in the determination of the Hero figure. From ISB responses obtained in the pilot study, as well as from a large pool of items gathered previously by others, a manual for scoring ISB responses for n Dependency was developed in a similar fashion. In making their ratings, the judges were asked to use the manuals by comparing the TAT or ISB item to be rated with examples included in the manuals. Approximately 30 hours were spent in training the judges in the use of the manuals.

Procedure

The *S*s were first seen at their respective fraternity houses, where the sociometric and projective measures were obtained. Following the group's completion of the sociometric ratings, the TAT cards were projected on a screen. *S*s were allowed six minutes, the TAT card remaining in view, to write each story. Finally, the ISB was administered. Each *S* was interviewed individually within a two-week period following the group testing. All interviews were recorded on the Gray Audograph, and the judges made their ratings of the interviews from these recordings. The experimenter and two interview judges, working independently, rated the interviews. The experimenter and two different judges, working independently, rated the projective test responses. The protocols and interviews were coded so that the identity of the individual *S*s was not known to the judges. All TAT and ISB responses for any given card or item were rated before continuing to the next item, and all responses were rated by at least two judges.

Estimates of the interrater reliabilities for the interview variables, the TAT, and the ISB were obtained by correlating (Pearson r) the ratings obtained from each judge with those of each of the other two judges. The interview variables were each rated on a ten-point scale. For the TAT and the ISB, each *S* was given a score by counting the number of stories or items rated as Dependent. The average interrater correlations for the interview were as follows: Dependent behavior, .54; Need Value, .45; Freedom of Movement, .56. The average interrater correlations were for the TAT (Hero), .75; TAT (Other), .11; and the ISB, .78. Because of the unreliable TAT Other Figure ratings, this index was dropped from consideration. The sociometric ratings yielded corrected split-half reliability of .83.

RESULTS

A test of the first hypothesis concerning the relationship between projective test responses and overt Dependent behavior was made by correlating (Pearson r) the interview and the sociometric ratings of Dependent behavior with the frequency-of-response ratings on the

---

[3] The method of nominating only the extremes of a group has been found to be at least as reliable as rating an entire group and to provide a distribution of ratings similar to that obtained when each member of a group is to be rated by every other member.

TAT and on the ISB. It appears from Table 1 that there is no evidence of a relationship between frequency of response in n Dependency on the TAT and either the interview or the sociometric measures of Dependent behavior. The ISB, in terms of frequency of response in n Dependency, yields indices which are related to both at low but significant magnitudes.

The second hypothesis concerning the relationship of frequency of response in n Dependency to importance of conflict in n Dependency, as determined from the interview, was evaluated by the chi-square test. In order to make this test, *S*s were divided into two groups for each projective test respectively, i.e., those *S*s with a high number (four or more on the TAT; seven or more on the ISB) vs. those with a low number (three or less on the TAT; six or less on the ISB) of Dependency responses. The high and low groups for each test were compared on the basis of an importance-of-conflict score which was derived in the following manner: *S*s with Freedom of Movement equal to or greater than Need Value were considered to have "No Important Conflict" in n Dependency and were compared with those *S*s whose Freedom of Movement was less than Need Value. These latter *S*s were considered to have an "Important Conflict" in n Dependency. The results of this comparison are reported in Table 2. It appears that those *S*s who have a relatively higher Need Value than Freedom of Movement for Dependency satisfaction give significantly more responses in n Dependency on the TAT. The results for the ISB are not significant.

It seems evident also from Table 2 that the TAT tends to elicit dissimilar data, in terms of a frequency-of-response measure, from that elicited by the ISB. Comparisons of these two tests in terms of frequency of response in n Dependency as related to interview and sociometric measures of Dependent behavior (Table 1) support this conclusion. Further evidence that these two instruments tend to elicit dissimilar data is given by their low intercorrelation (.09).

TABLE 1

CORRELATIONS BETWEEN THE MEASURES OF DEPENDENT BEHAVIOR AND FREQUENCY OF RESPONSE RATINGS ON THE PROJECTIVE TESTS

($N = 60$)

	TAT	ISB
Interview	.07	.28*
Sociometric	.06	.25*

* $p = .05$.

TABLE 2

CHI SQUARE OF THE FREQUENCY OF DEPENDENCY RESPONSES ON THE PROJECTIVE MEASURES COMPARED TO CONFLICT IN NEED DEPENDENCY

($N = 60$)

Conflict	Total Number of Dependency Responses			
	TAT		ISB	
	High D Group	Low D Group	High D Group	Low D Group
"Important Conflict" (Freedom of Movement less than Need Value)	88	27	183	79
"Not Important Conflict" (Freedom of Movement equal to or greater than Need Value)	16	18	65	37
	$\chi^2 = 12.24$; $p = .001$.		$\chi^2 = .99$; $p = .30$.	

On the other hand, the correlation between ratings of overt Dependent behavior from the interview and from sociometric measures was .59. Important to note here is the potential usefulness of the economical sociometric technique in relation to the laborious clinical interview as a quantified index of certain kinds of overt behavior.

CONCLUSIONS AND DISCUSSION

The following conclusions, all in terms of frequency of response in n Dependency, may be drawn from this study:

1. There is no evidence that TAT responses are related to interview or sociometric indices of Dependent behavior.
2. TAT responses are significantly related to an interview measure of importance of conflict in n Dependency.

The foregoing conclusions lend support to two assumptions listed by Lindzey in his review of the clinical use of the TAT. One assumption states, "The dispositions and conflicts that may be inferred from the storyteller's creations are *not always* reflected *directly* in overt behavior." Another holds that, "Themes that are recurrent in a series of stories are particularly apt to mirror the impulses and conflicts of the story-teller" (4, p. 3). The similarity of these assumptions to the hypotheses in this study is evident.

3. ISB responses tend to be related to interview and sociometric measures of overt, Dependent behavior.

4. ISB responses show no evidence of relationship to an interview measure of importance of conflict in n Dependency.

The third conclusion is similar to that of McClelland (5), who, using a Sentence Completion Test to measure n Achievement, argued that this test provides results which are dissimilar to those provided by the TAT and which are more similar to conscious self-judgments. Blyth (1) found also that the ISB differentiated between groups of veterans who behaved dependently or independently in terms of accepting or rejecting offers of psychotherapy.

It is evident that the TAT and the ISB, although both projective tests, provide dissimilar data in terms of frequency of response in a given need category, and that inferences from one test cannot be generalized to the other. More specifically, where inferences about overt behavior are to be made, a highly structured, behaviorally oriented projective test, such as the ISB, may be the more appropriate. The TAT does not appear to be particularly useful as a source of inferences about overt behavior. Where inferences about conflict in a given need area are to be made, a projective test that allows considerable freedom of response, such as the TAT, is probably appropriate. The ISB appears to be less appropriate as a source of inferences about conflict in a given need.

The unmodified TAT and the ISB were used in this study as they generally are employed in clinical psychological work. It is important to emphasize again that the measure used in relation to the projective test responses was a frequency-of-response measure. It is not intended to imply that such a measure of projective test responses is the only possible one or that such a measure is more valid or economical than other possible uses of such projective test responses. In fact, it is evident that a frequency-of-response measure makes no more than minimal use of the data provided by the TAT and the ISB. Nevertheless, with descriptive constructs from Rotter's social learning theory, which provided for an objective and communicable method, a predictive relationship was demonstrated between frequency of response in n Dependency on the TAT and an interview measure of importance of conflict in n Dependency.

SUMMARY

Data relevant to two hypotheses concerned with the evaluation of projective test responses were obtained in this study. The projective test measure was frequency of response in n Dependency on the TAT and on the ISB. A structured clinical interview and a sociometric technique were used as measures of overt Dependent behavior. The clinical interview also was structured to provide measures of Need Value and Freedom of Movement for n Dependency. Eighty *S*s participated in the study. A pilot study with 20 of these *S*s provided data for judges' scoring manuals for the interview and projective test situations. All *S*s were tested in groups and interviewed individually.

Results of the study may be summarized as follows:

1. Frequency of response in n Dependency on the TAT yielded no evidence of relationship to the interview and sociometric measures of overt Dependent behavior.

2. Frequency of response in n Dependency on the ISB was related to interview and sociometric measures of overt Dependent behavior.

3. Importance of conflict in n Dependency, as derived from the interview, was related to frequency of response in n Dependency on the TAT. The results for the ISB were not significant.

4. Dissimilar data, in terms of frequency of response in n Dependency, were elicited by the TAT and the ISB.

5. A high degree of similarity was found between the sociometric and interview ratings of overt Dependent behavior. This suggests the usefulness of sociometric ratings in lieu of clinical interviews to obtain an approximate, quantifiable measure of certain kinds of overt behavior.

REFERENCES

1. BLYTH, D. D. Dependency, independency, and other factors related to veterans' reactions to an offer of psychotherapy. Unpublished doctoral dissertation, Ohio State Univer., 1953.
2. HANFMANN, E. William Stern on "Projective Techniques." *J. Pers.*, 1952, **21**, 1–21.
3. KORNER, A. F. Theoretical considerations concerning the scope and limitations of projective techniques. *J. abnorm. soc. Psychol.*, 1950, **45**, 619–627.
4. LINDZEY, G. Thematic Apperception Test: Interpretative assumptions and related empirical evidence. *Psychol. Bull.*, 1952, **49**, 1–25.
5. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
6. ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
7. SOSKIN, W. F. Bias in postdiction from projective tests. *J. abnorm. soc. Psychol.*, 1954, **49**, 69–74.

Received February 11, 1957.

THE REPRESSION HYPOTHESIS STUDIED IN A SITUATION OF HYPNOTICALLY INDUCED CONFLICT

RUTH A. BOBBITT[1]

State University of Iowa

In recent years, a number of experimental psychologists have tried to rewrite psychoanalytic theory in operational terms. The concept of repression has stimulated a variety of studies (13, 16, 21, 22, 23, 24, 25) that have succeeded primarily in pointing up methodological problems but have contributed little to our knowledge of repression as either a phenomenon or a process. This paper reports a study made ten years ago (3) that still seems pertinent today.

Repression, as a method of ego defense, protects the individual from painful emotional states by keeping ideas that represent certain impulses out of consciousness (7, p. 86). Following Freud's formulation, we may identify the necessary conditions for the development of repression as these: (*a*) An instinctual impulse is aroused; (*b*) gratification of this impulse is in some way unacceptable to the ego of the individual; (*c*) pain (anxiety) has been experienced as the ego demands have been frustrated in gratifying the instinctual tendency (7, p. 84; 9, p. 119); and (*d*) the need to avoid this anxiety is stronger than the instinctual drive (7, p. 85).

As a process, repression as conceived by Freud has the following dynamic characteristics:

1. The repressed is only separated from consciousness (awareness) and is presumably still active in the unconscious. In fact, the extent and degree of repression varies as inhibitory and instinctual propensities fluctuate (7, pp. 89–90).

2. Generalization of cues and responses often occurs. Cues originally related to the drive may be repressed, but other representations may be admitted to consciousness and acted upon with the accumulated energy of the repressed (7, pp. 86–87). Or, the related ideas may become so associated with the repressed that they must also undergo repression if the individual is to avoid anxiety (Freud's "repression proper," (7, p. 86). If repressed impulses are satisfied, it must be through substitute activities. Unless such satisfactions are complete, the need persists, to be expressed through an increasing number of substitute ideas and with "the damming-up consequent on lack of real satisfaction" (7, p. 87).

Since "the motive and purpose of repression is simply the avoidance of 'pain' " (7, p. 92), it follows that repression may be considered successful only to the extent that it does avoid anxiety or pain. It has been noted by Freud (10, pp. 21–22) and by Merrill (16) and Zeller (23) that unsuccessful, i.e., partial repression, or repression in which at least the affective component has become attached to a substitute idea is most amenable to analysis or experiment. These emotional components become apparent when one considers the dynamics of psychoanalytic therapy in which the aim is to remove repressions. Freud says, ". . . the patient can go on spinning a whole chain of . . . associations, till he is brought up in the midst of them against some thought-formation, the relation of which to what is repressed acts so intensely that he is compelled to repeat his attempt at repression" (7, p. 88). Finally, summarizing the formula for therapy, ". . . all repressions are to be undone; the mental condition is then the same as if all amnesias are removed. When all gaps in memory have been filled in, all the enigmatic products of mental life elucidated, the continuance and even the renewal of the morbid condition is impossible" (8, p. 269).

In this theoretical context, this investigation is specifically concerned with anxiety behavior as the response variable to be measured as a function of degree (completeness) of repression. In terms of the general hypothesis, and apart from therapeutic intervention, the following relationships are predicted: In the absence of either conflict or repression, there should be no anxiety behavior. In conflict with full repression, there should be no anxiety behavior except as the repressive process fluctuates. With incomplete repression, there should be an increase in the scope and intensity of anxiety disturbance. As repression ceases to operate disturbance should reach a maximum and show considerable generalization. Finally, when both repression and conflict are removed there should no longer be anxiety behavior related to the conflict.

---

[1] The investigation was conducted under the direction of Professor Charles R. Strother; technical assistance was given by Drs. Wilbur R. Miller and Paul Huston of the Psychopathic Hospital, and by Professors Robert R. Sears and E. E. Lindquist of the State University of Iowa. The author now resides at 3900–50 N.E., Seattle, Washington.

Most of the investigations of repression have utilized some measures of memory as the response variable. Zeller (23, pp. 40–45) recently provided an excellent critical review of these, pointing out the fallacies and inadequacies of each. The kind and amount of feeling tone, emotional disturbance, or ego threat in the experimental situation are seen as important factors. He also stresses that any critical study must demonstrate "restoration to consciousness of the repressed material" upon removal of the repression factor (23, p. 46).

Zeller (23, p. 44) and other writers (16, 22) have noted the Huston, Shakow, and Erickson experiment (13) as one of the better demonstrations of repression, although the use of hypnosis makes interpretation somewhat difficult. The criticism in terms of dependence on hypnosis loses some weight when we consider that results in conformity with the repression hypothesis occur as predicted, and that these results could not have been predicted on the basis of hypnosis alone. The hypnotic approach surely warrants further investigation, if only for the fact that other approaches thus far devised involve variables perhaps even more difficult of appraisal, i.e., learning, motivation, reminiscence, retroactive inhibition, etc. (24, p. 422).

A further consideration in experimental design is the response to be measured as a function of repression. In the case of most of the experiments measuring memory responses, "anxiety" or some form of personal involvement (frustration, incompleteness, unpleasantness) is inferred. In the study of Huston, Shakow, and Erickson (13), however, the responses measured are presumed on the basis of tests of validity to be measures of emotional disturbance or anxiety. The condition of repression is of course also independently validated.

METHOD

This experiment employs hypnotically induced conflicts as described by Erickson (6) tested according to Luria's association test procedures (13, 15). Experimentally induced conflicts offer opportunity for necessary control. Except when hypnosis has been employed, however, studies of conflict in human subjects (*S*s) have been too circumscribed and moderate for our purposes. The hypnotic method has been found to have some validity, i.e., the experimenter can define the nature of the conflict, assure the *S*'s susceptibility to and acceptance of it, suggest in general the nature of his reaction to it, create the effects of repression and, finally, remove the hypnotic effects without undesirable residual reactions.

Luria's method for testing emotional disturbance calls for conditioning a voluntary motor pressure of the preferred hand to occur simultaneously with a verbal response. The theory is that an integrated response of low habit strength under voluntary control is notably susceptible to disorganizing effects of emotional disturbance. His report of the general validity of the method has been verified in a group of studies (1, 4, 5, 11, 12, 13, 18, 19, 20).

The design of this experiment is as follows: Under hypnosis, an incident created to arouse a severe state of conflict is related to six selected *S*s. In the waking state, responses to a standard set of 100 association test words are recorded under five experimental conditions: (*a*) before the conflict is induced; (*b*) after induction of the conflict and with posthypnotic amnesia for it; (*c*) with partial posthypnotic amnesia; (*d*) with full awareness of the conflict as an immediate actuality; and (*e*) after hypnotic removal of the conflict. The association test is composed of "critical" words relating directly to the conflict story; "postcritical" words immediately following critical words, and "neutral" words to which the *S*'s responses were in no way disturbed on a preliminary testing. The three experimental variables are thus subjects (assumed to contain a random effect), types of verbal stimuli and conditions of conflict awareness. The simple factorial design permits evaluation of the three variables in all combinations by means of analyses of variance.

Subjects

*S*s were selected from a group of male junior and senior medical students who had no previous experience with hypnosis, but were willing to serve in a reaction time experiment involving hypnosis. After training, each *S* was rated hypnotizable and capable of complete posthypnotic amnesia.[2] No individuals were accepted whose personalities were rated "extreme" on the basis of a brief Rorschach evaluation, a psychiatric interview and behavior during hypnotic training sessions.

Procedure

The same conflict story and instructions governing degree of awareness were related under hypnosis to each *S*. In outline, the story was this:

In medical school, the *S* has had an experience so disturbing that he had to forget it. As it is recalled, he will re-experience it in detail. He was attending a party and volunteered to borrow a car and go for more beer. On his way back, he drove rapidly around a dark corner and felt the car hit something. He feared an accident but did not want to stop since he was in a borrowed car and had been drinking. He rationalized that it had probably been only a bump in the road and drove on. Next morning his fears were confirmed in a newspaper story of a child, near death and hopelessly crippled as the result of a hit and run accident. The details of the account made it clear to him that he was the driver for whom the police were searching. He experienced great panic and conflict about what to do.

Each condition of awareness was induced under hypnosis after review of the story. For the postconflict condition, the hypnotist proved to *S* under hypnosis that the incident could not have occurred and talked freely with him to assure removal of the conflict. After completion of the experiment, a final hypnotic session assured removal of the conflict; it was suggested also that the *S* would recall his thoughts and feelings, wanting to talk freely about them when he awakened.

The association test was administered before hypnosis (Condition I), in the waking state after each hypnotic session (Conditions II, III, and IV), and after the conflict was removed (Condition V).

Independent ratings of each *S*'s behavior and introspective accounts were made by two judges, the hypnotist and the *E*. Unanimous agreement that *S* had accepted the conflict and with respect to levels of awareness was required.

The 100-word association test, composed of 60 neutral, 20 critical, and 20 postcritical words, was individually constructed for each *S*. Neutral and postcritical words were selected on the basis of *S*'s responses to a preliminary test of 210 words. Words were excluded when any measure of response deviated markedly from his mode. Eighty words were then drawn at random from the remaining list. The critical words were either key words in the conflict story (*collision, party, newspaper, borrow,* etc.) or suggested feelings and ideas closely related to it (*criminal, guilty, conflict,* etc.). These words were randomized and distributed at varying intervals among the neutral words. The same order of words was maintained throughout the testing.[3]

The apparatus employed (2, 3) furnished several response measures: verbal reaction time (VRT), time between beginning of verbal response and beginning of voluntary motor pressure (VR-VMP), number of VMP made during a single stimulus-response sequence (#VMP), duration of the first VMP in a stimulus-response sequence (DVMP), number of involuntary motor pressures occurring in a stimulus-response sequence (#IMP), duration of IMP (DIMP), breathing rate (BR), and number of conflict-associated verbal responses obtained during a single testing period (#CAVR). These measures were reliable, as indicated by agreement between two scorers. Reliability coefficients range from .84 to .997 on measured items and percentages of agreement range from .916 to 1.00 on counted items.

RESULTS

The validity of the experimental situation, apart from that of the response measures, is assumed on the basis of unanimous judges' ratings and on the consistency among observations and introspective reports. Furthermore, there appears to be descriptive similarity between repression, as postulated here, and the posthypnotic conditions of awareness.

Validity of response measures could be tested more directly. By means of analyses of variance, the three experimental variables (subjects, stimuli, and conditions) were tested in interaction with one another (14). Since none of these tests was significant, the specific differences relating to validity and to the experimental hypotheses were tested by the triple interaction variance as the error term.

Table 1 summarizes results of the validity tests.[4] It is seen that (*a*) all response measures, except BR, tend to differentiate responses in the conditions of conflict from those in conditions of nonconflict (see Lines 1*a*, 1*b*, and 1*c*); (*b*) all measures but DVMP tend to differentiate between responses to critical and to

---

[2] Hypnosis was conducted by three psychiatrists, Drs. Paul Huston, A. W. Freidinger, and F. S. Bobbitt.

[3] Maintaining the same sequence of words throughout the experiment makes likely the operation of practice effects in opposition to the experimental hypotheses. Changing sequence, however, might introduce factors less subject to appraisal.

[4] Tests significant at or below the .05 level of confidence are considered "significant"; those between the .05 and .20 levels are considered "suggestive." This procedure is used to point up consistent trends. The small number of *S*s makes necessary large differences, considerable precision and intersubject homogeneity for a high level of statistical significance; requirements difficult of fulfillment in an experiment of this kind.

TABLE 1

RESULTS OF VALIDITY TESTS OF ALL RESPONSE MEASURES SHOWING FOR EACH THE LEVEL OF CONFIDENCE WITH WHICH THE HYPOTHESIS OF NO DIFFERENCE MAY BE REJECTED

Difference tested	Response Measure[a]							
	VRT	VR-VMP	DVMP	# VMP	BR	# IMP	DIMP	CAVR
1. Between means for conflict and nonconflict conditions								
a. Critical words, conflict minus critical words, nonconflict	1%	2–5%	5–10%	10–20%	—[b]	5–10%	5–10%	—[b]
b. All words, conflict minus all words, nonconflict					20%			
c. All words, conflict minus all words, postconflict								10–20%
2. Between mean response to critical and to neutral stimulus words								
a. Critical words, conflict minus neutral words, conflict	1%	5–10%	20%	10–20%	—[b]	10–20%	10–20%	—[b]
b. Critical words, nonconflict minus neutral words, nonconflict	20%	20%	20%	20%	—[b]	20%	20%	—[b]

[a] Response measure:
VRT—Verbal reaction time
VR-VMP—Time between verbal response and voluntary motor pressure
DVMP—Duration voluntary pressure
VMP—Voluntary motor pressure
BR—Breathing rate
IMP—Involuntary motor pressure
DIMP—Duration involuntary pressure
CAVR—Conflict associated verbal response

[b] Response measure could not be tested for this difference.

neutral words under the conflict conditions (Line 2*a*). Conversely, no difference between responses to these words is shown in the nonconflict conditions (Line 2*b*). It is interesting to note, in terms of Luria's findings, that VRT and VR-VMP can be considered valid with greater confidence than any of the other measures.

It is now possible to evaluate the findings in relation to the experimental hypotheses. When hypothetical results (Fig. 1) are compared with obtained results for the group (Fig. 2), general conformity is apparent. However, maximum disturbance tends to occur more often in Condition III (partial awareness) than in Condition IV (full awareness), and there is little generalization of disturbance in responses to postcritical and neutral stimuli.

The specific experimental predictions have been subjected to statistical test; reported in Table 2, however, are only those tests warranted by the validity tests (Table 1) and study of Fig. 2. Breathing rate and conflict-associated verbal responses are omitted as are comparisons involving responses to postcritical and neutral words.

1. There was little disturbance under the condition of conflict without awareness (Table 2, Line 1). Only VRT and DIMP tend to differentiate between these conditions. That there is no particular disturbance in relation to other types of stimuli is evident in Fig. 2.

2. There is increased disturbance in response to critical words in the condition of partial awareness over the condition of conflict without awareness (Table 2, Line 2) and over the condition of preconflict (Table 2, Line 3). There is also the predicted involvement of more response measures (Table 2, Line 3 in relation to Line 1), but obviously not a significant generalization of disturbance to postcritical and neutral stimuli (Fig. 2).

3. The predicted further increase (to a maximum) in disturbance under the condition of conflict with full awareness (Table 2, Lines 4 and 5) did not occur. Fig. 2 shows that, for most measures, disturbance in this condition (IV) is somewhere between the other two conflict conditions. For all measures though, responses to critical words are significantly different in the condition of full awareness from those in the preconflict condition (Table 2, Line 6).

4. All response measures lose their disturbed character when conflict is removed but awareness maintained (Table 2, Line 7). No response

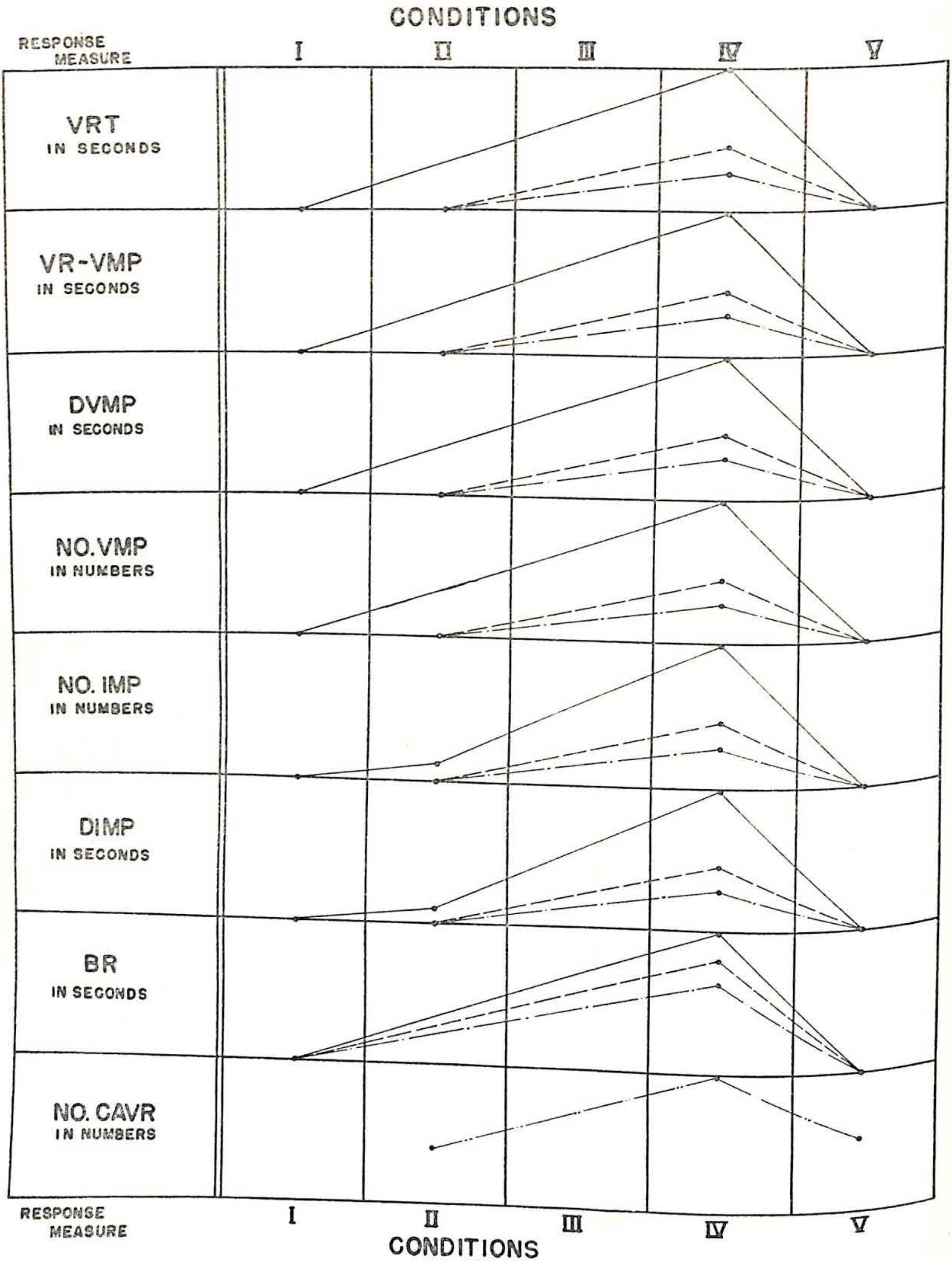


FIG. 1. HYPOTHETICAL CURVES OF RESPONSE MEASURES FOR THE TOTAL GROUP BY EXPERIMENTAL CONDITIONS AND TYPES OF STIMULI

LEGEND:
CRITICAL ————
POST CRITICAL – – – –
NEUTRAL — · —

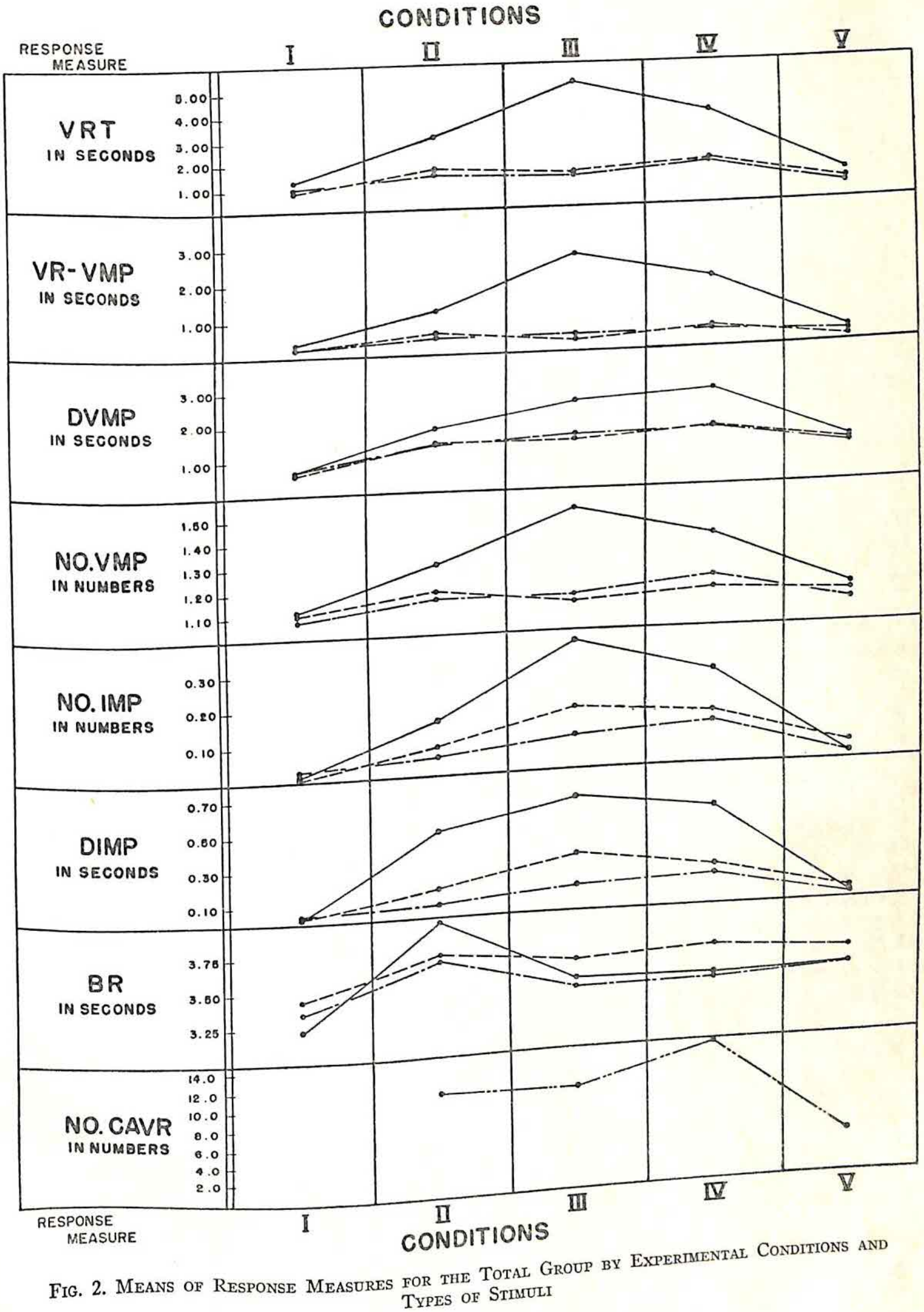


FIG. 2. MEANS OF RESPONSE MEASURES FOR THE TOTAL GROUP BY EXPERIMENTAL CONDITIONS AND TYPES OF STIMULI

LEGEND:
CRITICAL ————
POST CRITICAL — — — —
NEUTRAL — - — - —

TABLE 2

TESTS OF THE "DEGREE OF AWARENESS" HYPOTHESIS SHOWING THE LEVEL OF CONFIDENCE WITH WHICH THE HYPOTHESIS OF NO DIFFERENCE MAY BE REJECTED
(For each response measure, for critical stimulus words only)

Difference Between Means	Response Measure					
	VRT	VR-VMP	DVMP	VMP	IMP	DIMP
1. Preconflict minus conflict without awareness	5–10%	20%	20%	20%	20%	10–20%
2. Conflict without awareness minus conflict with partial awareness	1–2%	5–10%	20%	10–20%	10–20%	20%
3. Preconflict minus conflict with partial awareness	1%	1%	2–5%	1–2%	1%	2–5%
4. Conflict with partial awareness minus conflict with full awareness	10–20%[a]	20%	20%	20%	20%	20%
5. Conflict without awareness minus conflict with full awareness	20%	20%	10–20%	20%	20%	20%
6. Preconflict minus conflict with full awareness	1%	2–5%	1–2%	10–20%	2–5%	5–10%
7. Preconflict minus postconflict	20%	20%	20%	20%	20%	20%

[a] This tendency is for verbal reaction time to be significantly less disturbed.

measure differs significantly between preconflict and postconflict conditions. With no data available on conflict-associated verbal responses in the preconflict condition, it is not possible to evaluate the significance of the average of four such responses in the postconflict condition (Fig. 2).

DISCUSSION

Two discrepancies, one major and one minor, between the experimental hypotheses and obtained data have been noted. First, maximum disturbance did not occur in Condition IV. In this connection, it must be considered that:

1. All obtained differences between conditions of awareness may be due to chance. However, the consistent statistical, graphic, and introspective results make this a remote possibility.

2. Contrary to the judges' evaluations and the introspective accounts, there may have been no real difference in awareness between Conditions III and IV.

3. With development of full awareness, the situation may have had less conflict-value. No *S* expressed doubt that the accident occurred. Three *S*s stated that they feared inevitable punishment and show a general downward trend in disturbance. Two *S*s had decided against conscience and in favor of self-protection. One of these showed greatest disturbance in Condition IV, while the other showed somewhat inconsistent increases and decreases in disturbance. One *S*, whose behavior was much the same in Conditions III and IV, could see no solution to the conflict but was trying to avoid detection.

If there was resolution of conflict in Condition IV, the hypothesis as formulated may be true, but most of the test may have been administered after resolution began and may measure the process of conflict and anxiety reduction instead of reaction to the development of full awareness.

4. The data may reflect an error in the formulation of the hypothesis concerning degree of awareness. The development of full awareness of a conflict may be accompanied by diminution in disturbance so that maximum disturbance occurs when awareness is at least partial but somewhat less than complete. The peak of disturbance may come with developing awareness, but almost immediately other defenses operate to reduce disturbance (anxiety).

5. The effects of full awareness may have been counteracted by some habituation factor. Tests of practice effects, i.e., differences between Conditions I and V, do not show significant results (Table 2, Line 7). Tests of abreaction or desensitization cannot be made separate from those of the experimental hypotheses.

The second discrepancy is that the expected generalization of disturbance to postcritical and neutral stimulus words did not materialize to any significant extent. The qualitative findings of Luria and of Huston, Shakow, and Erickson (13, 15) suggest such generalization, while the statistical findings of Gardner (11) do not. Generalization may depend on several factors, two of which should be mentioned here:

1. The time interval between a verbal response and the succeeding verbal stimulus (in this study, approximately three seconds were allowed to elapse to permit dissipation of disturbance in the appropriate stimulus-response sequence).

2. The remoteness of postcritical and neutral words from critical subject matter (in this experiment, rather unusual care was taken, in selecting neutral words, to avoid critical implications).

SUMMARY

Based on the Freudian concept of repression, an experiment was designed to test the following specific hypotheses with respect to anxiety behavior:

1. Repression serves to protect the individual from anxiety related to the repressed impulse.

2. The less successful the repression, the greater is the anxiety experienced in relation to conflict between the repressed and the ego demands.

3. Complete failure of repression results in maximum anxiety, a condition which persists until a new repression is instituted or another defense becomes operative.

4. Complete relief from both repression and conflict results in cessation of anxiety.

Through hypnotic induction and control of conflict and degree of awareness, five experimental conditions were produced: (*a*) preconflict control; (*b*) conflict with posthypnotic amnesia for the situation; (*c*) conflict with partial posthypnotic amnesia; (*d*) conflict with full awareness; and (*e*) postconflict control. Eight measures of response to a word-association test administered by the Luria method were obtained for six *S*s. Each *S* was tested in the waking state under the five experimental conditions.

For the *S*s as a group it was demonstrated that (*a*) there was a conflict; (*b*) there was a progressive increase in degree of awareness; and (*c*) most of the measures of behavior reflected emotional disturbance with some validity.

Results were as predicted except in two respects:

1. There was not the anticipated maximum disturbance when *S*s were made fully aware of the conflict. It is suggested that (*a*) with failure of repression certain defenses quickly operate to diminish anxiety, or (*b*) the hypothesis requires reformulation at this point.

2. Predicted generalization of disturbance to neutral stimulus material did not occur, although there was generalization to an increasing number of response measures. It appears that the testing procedure allowed most disturbances to dissipate in the stimulus-response sequence in which they began.

REFERENCES

1. BARNACLE, C. H., EBAUGH, P. G., & LEMERE, F. Association motor investigation of the psychoneuroses. *Amer. J. Psychiat.*, 1935, **91,** 925–937.
2. BOBBITT, RUTH A. A new apparatus for the Luria experiment. *J. exp. Psychol.*, 1940, **27,** 578–582.
3. BOBBITT, RUTH A. An experimental investigation of certain aspects of the repression hypothesis. Unpublished doctoral dissertation, State Univer. of Iowa Library, 1947.
4. BURTT, H. E. Motor concomitants of the association reaction. *J. exp. Psychol.*, 1936, **19,** 51–63.
5. CROSLAND, H. R. The psychological methods of word-association and reaction-time as tests of deception. *Pub. Univer. Oregon, Psychol. Series I*, 1929, **1,** 104.
6. ERICKSON, M. H. The method employed to formulate a complex story for the induction of an experimental neurosis in a hypnotic subject. *J. gen. Psychol.*, 1944, **31,** 67–84.
7. FREUD, S. Repression. *Collected Papers*, 1915, **4,** 84–97.
8. FREUD, S. Early papers and history of psychoanalytical movement. *Collected Papers*, 1924, **1,** 379.
9. FREUD, S. *New introductory lectures in psychoanalysis.* New York: W. W. Norton, 1933.
10. FREUD, S. *The problem of anxiety.* New York: Psychoanalytic Quarterly Press & W. W. Norton, 1936.
11. GARDNER, J. W. An experimental study of the Luria technique for detecting mental conflicts. *J. exp. Psychol.*, 1937, **20,** 495–506.
12. HOUTCHENS, H. M. A study of mental conflict in delinquent and nondelinquent boys. Master's thesis, Univer. of Iowa, 1935. (Abridged in *J. Juv. Res.*, 1935, **19,** 180–192.)
13. HUSTON, P. E., SHAKOW, D., & ERICKSON, M. H. A study of hypnotically induced complexes by means of the Luria technique. *J. gen. Psychol.*, 1934, **11,** 65–97.

14. Lindquist, M. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.

15. Luria, A. *The nature of human conflicts.* (Trans. by W. Horsley Gantt.) New York: Liveright, 1932.

16. Merrill, R. An experimental study of repression. Unpublished doctoral dissertation, Univer. of Washington Library, 1953.

17. Miller, N. E., & Dollard, J. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

18. Morgan, M. I., & Ojemann, R. H. A study of the Luria method. *J. appl. Psychol.*, 1942, **26,** 168–179.

19. Olson, D. M., & Jones, V. An objective measure of emotionally toned attitudes. *J. genet. Psychol.*, 1931, **39,** 174–196.

20. Reymert, M. L., & Speer, G. S. Does the Luria technique measure emotion or merely bodily tension? *Character & Pers.*, 1939, **7,** 192–200.

21. Rosenzweig, S. Investigation of repression as an instance of experimental idiodynamics. *Psychol. Rev.*, 1952, **59,** 339–345.

22. Sears, R. R. Functional abnormalities of memory with special reference to amnesia. *Psychol. Bull.*, 1936, **33,** 229–274.

23. Zeller, A. F. An experimental analogue of repression. I. Historical summary. *Psychol. Bull.*, 1950, **47,** 39–51.

24. Zeller, A. F. An experimental analogue of repression. II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, **40**, 411–422.

25. Zeller, A. F. An experimental analogue of repression. III. Effect of induced failure and success on memory measured by recall. *J. exp. Psychol.*, 1951, **42,** 32–38.

Received April 22, 1957.

THE EFFECT OF SIBLING COMPETITION AND ALLIANCE ON LEVEL OF ASPIRATION, EXPECTATION, AND PERFORMANCE[1]

EVA DREIKURS FERGUSON

Western Psychiatric Institute, University of Pittsburgh

The present study seeks to test certain hypotheses that have arisen from the Adlerian theory of sibling relationships. According to this theory, personality patterns are a reflection of goals that the individual establishes within the family constellation of the early years of childhood. A goal basic to all human behavior is the wish to belong, to be accepted (1, 2, 4), so that the child's behavior patterns are expressions of trying to find a place, or status, within the family group. The presence of siblings in the family can be viewed by the child as an obstacle to attaining status, in which case the relationship he establishes with his siblings is a competitive one. If, however, the sibling is viewed as an ally in the process of gaining status in the family, the interaction between these two children takes on a pattern different from that of the competitive relationship. Since early childhood strivings for status find expression in individual personality development, the personality patterns of the siblings reveal the type of relationship established between them: allies become alike in personality, competitors become different. Dreikurs, in expanding the original conceptualization of Adler, has described as follows the dynamics of the sibling interaction and their effects on personality development:

> As each member of the family strives for his own place within the group, the competing opponents watch each other carefully to see the ways and means by which the opponent succeeds or fails. Where one succeeds, the other gives up; where one shows weakness or deficiencies, the other steps in. In this way competition between two members of the family is always expressed through differences in character, temperament, interests, and abilities. Conversely, the similarity of characteristics always indicates alliances. Sometimes, the two strongest competitors show no sign of open rivalry but, rather, present a close-knit pair; nevertheless, their competitive striving is expressed in personality differences (p. 40, 4).

Competition, as used here, thus describes a specific behavior pattern: giving up where the other person succeeds and striving to outdo only where the other fails, with the result that competitors display contrasting personality patterns. Dreikurs differentiates rivalry from competition (4, 5), rivalry being a struggle for superiority in specific situations or for limited goals, while competition is a pervasive and long-range struggle over the attainment of status within the family group. Moreover, rivalry need not display the "seesaw" phenomenon postulated above for competition. According to this distinction, rivalry may occur between any two individuals, including allies.

This conceptualization of competition and its effect on personality development has found validation in clinical practice, as for example in therapy with "problem children": once the problem child improves with therapy, his previously "good" competitor usually displays problem behavior that is entirely novel for him. Rigorous experimental tests of these formulations have not yet been conducted, however. The present study therefore attempted to clarify to what extent siblings who are Allies and those who are Competitors, as defined by sameness or oppositeness of traits, do behave in the manner described by Dreikurs. For this purpose, sibling pairs were given three experimental tasks, and two types of response measures were utilized: (*a*) performance and (*b*) statements of aspiration. Two hypotheses were derived from Dreikurs' position:

Hypothesis 1. In a competitive relationship, a child's performance, level of aspiration, and level of expectation *rise* when he is informed that his sibling has done more poorly than himself and *drop* when he is told that his sibling has done better than himself.

Hypothesis 2. In an alliance relationship, a child's performance, level of aspiration, and level of expectation *drop* when he is informed that his sibling has done more poorly than himself and *rise* when he is told his sibling has done better than himself.

---

[1] This study is a condensation of the doctoral dissertation (6) conducted in partial fulfillment for the Ph.D. degree at the Department of Psychology, Northwestern University. Grateful acknowledgments are extended to Drs. Benton J. Underwood, dissertation supervisor, and Rudolf Dreikurs, who served as interviewer and gave invaluable suggestions.

On the basis of data to be presented, it was found that these hypotheses required substantial modification. Differences did emerge between Competitors and Allies, but not as hypothesized nor did they occur on all tasks and response measures. Upon *post hoc* consideration of the task dimensions in the present study, it seemed necessary to take into account two additional variables, ego-involvement and status-effect, in framing a more adequate test of the Adler-Dreikurs formulation. These variables are here so defined as to link on a common dimension ego-involvement and success-failure evocation, variables which in the literature are largely treated as discrete phenomena and which to the present writer appear conceptually related. *Ego-involvement* is here defined as occurring when *S* perceives the task or situation as important to the maintenance or raising of his status, either because of the nature of the task or because of instructions given to *S*.[2] *Status-effect* is defined as occurring when *S* experiences a *shift* in status. Thus, under conditions of failure *S* experiences a decrease of status, and under conditions of success *S* experiences an increase of status. Presumably, ego-involvement (calling status into play) heightens status-effect, although operationally ego-involvement and status-effect are differentiable. Both for the purpose of the present study, as well as for a clarification of these problems in the literature, this kind of distinction may be found useful.

Although ego-involvement and status-effect were not varied systematically in the present study, *post hoc* analysis of the experimental procedure and of the obtained data suggested a revised formulation that gives prominence to these two variables. This formulation is presented at this point to facilitate the exposition of results along lines believed by the investigator to be promising for further research. A basic hypothesis in the revised conceptualization is that differences between Allies and Competitors appear in a sibling comparison situation when ego-involvement and status-effect are present to a high or moderate degree but not when they are minimally present. When these variables are present to a *high* degree, as they are in much of the real-life experience of the child, Allies and Competitors may be expected to behave as predicted by the original hypotheses, i.e., Allies strive to do like the sib, Competitors give up when the sib does better and only strive to outdo the sib when the sib does worse. When ego-involvement and status-effect are *moderate*, however, Competitors behave in a manner analogous to rivalry while Allies are relatively unaffected by the sib comparisons, being more attentive to their own than to their sib's performance. This revised hypothesis leads to several ramifications. For one thing, it suggests that the alliance relationship is a weaker one than that of competition, stronger ego-involvement and status-effect being necessary for the Allies than for the Competitors to show a change in a situation of sib comparison. It further suggests that the type of behavior shown by Competitors is contingent upon whether the sib comparisons are presented in a context of moderate or of high ego-involvement and status-effect. Finally, the revised formulation does not predict difference between Allies and Competitors when presented with sib comparisons under conditions of *low* ego-involvement and status-effect, even though both groups may use knowledge of the sib's performance as cues for their own performance when they lack prior experience with the task. Goal-setting statements, possibly tapping motivational variables more directly than do performance measures, should show response changes more clearly than performance measures in the sib comparison situation. It may therefore be hypothesized that whereas performance and goal-setting statements differentiate the behavior of Allies and Competitors under conditions of high ego-involvement and status-effect, only goal-setting statements do so under moderate ego-involvement and status-effect.

METHOD

Subjects

Two duplicate experiments were performed. Center *Ss* came from families where at least one sib was brought, within the past 18 months, to a child guidance center[3] as a behavior problem; Normal *Ss* were selected from churches in the community in which the university is located. Each experimental group was composed of 20 sib pairs, 10 of Allies and 10 of Competitors, giving a total of 80 *S*s or 40 sib pairs. All *S*s came from families that had three or more children, *S*s being between 4½ and 10 years of age. Allies and Competitors were matched by pairs for age and where possible for sex, approximately half the *S*s in each group being under eight years of age and half being eight years or older. The *S*s were selected on the basis of information obtained in a brief interview with the mother of each pair of sibs. The interviewer was a prominent Adlerian psychiatrist who was instructed to perform the interviews in the manner usual in his psychiatric practice. His diagnoses based on the interview data were found, on analysis subsequent to the completion of the study, to follow a quantitative system, although he was not aware of this and followed what he considered was an intuitive approach in arriving at the diagnoses.

---

[2] This definition of ego-involvement covers more aspects than are usually described in the literature (e.g., 7, 8), and as such appears to have more face validity. It includes raising of status, rather than merely threat to the maintenance of status, and it defines status both in terms of self-esteem and social-esteem, for want of a satisfyingly clear delineation between the two types of esteem.

[3] The clinics are part of the Community Child Guidance Centers of Chicago, a nonprofit organization under the sponsorship of the Individual Psychology Association of Chicago.

A pre-experimental test of the reliability of the interview method was performed. In this, the interviewer talked to twenty sets of parents without knowledge of which father and mother belonged together. To assess degree of agreement between the diagnoses derived from interview with the father and from interview with the mother, a 3 × 3 chi-square contingency table was prepared for the three diagnostic categories of Allies, Mixed, and Competitors. This yielded a contingency coefficient of .640, significant beyond the .01 level. (This coefficient is to be compared to the maximum possible for a 3 × 3 table, of .816.) In the interviews, the examiner first asked the parent which of the children are most different and which most alike, and in what regard. He then asked the parent to describe each child. If more information was needed, the interviewer asked about the following types of traits: who has the most (least) temper; who minds best (least); who is the best (worst) student; who is the most (least) outgoing.

For the 80 *S*s used in the experiment, *E* did not know the diagnosis of *S*s until the completion of the study, at which time the interview material was analyzed. This analysis revealed that when sibs showed a ratio of at least three traits opposite to one trait alike, the diagnosis was that of competition, and when the trait ratio of alike to opposite was at least three to one, the diagnosis was that of alliance.[4] The 40 experimental pairs were drawn from a total of 104 diagnosed pairs, of which 10 pairs were not usable because one or both of the sibs were outside the required age range. The diagnostic distribution of the 104 pairs was 54 pairs of Competitors, 30 pairs of Allies, and 20 pairs of Mixed diagnosis. Selection of the final sample of 40 pairs was dictated by the requirements of matching Allies and Competitors on a pair basis for age and where possible, for sex. Table 1 presents relevant characteristics of *S*s.

TABLE 1

CHARACTERISTICS OF SIBLING PAIRS

Pair	Normal Sample		Center Sample	
	Competitors	Allies	Competitors	Allies
Number of Pairs				
$F_1 - F_2$[a]	4	3	1	5
$F_1 - M_2$	2	3	2	1
$M_1 - F_2$	2	2	3	1
$M_1 - M_2$	2	2	4	3
Total	10	10	10	10
Years[b]				
Mean age	7.23	7.20	7.60	7.68
Mean age difference	3.40	3.75	2.60	2.95
Mean Ratio				
Mean ratio of similar to opposite traits	.67/4.20	4.08/.35	.07/3.47	4.13/.57

[a] $F_1 - F_2$ is to be read that both older and younger siblings are female; $F_1 - M_2$, that the older sibling is female, the younger, male; etc.

[b] Years are computed in half-year units.

Tasks

Two aspiration tasks and one verbal learning task were administered. Both *S*s of the sib pair came to the laboratory at the same time, each *S* being put in a separate room. Each task was presented alternately to each sib, and *S*s had no contact with one another throughout the experiment. The only information each sib had of the other's performance was that supplied by *E*. While *E* tested one *S*, the sib played with an assistant *E*. in another room. Time intervals between tasks were arranged to control for warm-up and transfer effects and to make credible the information given to *S* regarding the sib's performance. To ensure *S*s' cooperation throughout the duration of the experiment, all *S*s were told prior to entering the laboratory that they would play some games with *E*, and for being so nice and coming to play the games *S* would get a small prize afterwards. So as not to interfere with the experimental conditions (i.e., the sib comparisons made to *S*) only a minimal amount of encouragement was given to *S*s throughout the experiment.

The first task presented to each *S* (Aspiration Task A) consisted of 40 pairs of dots. The *E* demonstrated how the dots could be connected, and then told *S* that within one minute most children of his and his sib's age could finish putting lines through the dots for 20 pairs of dots, but that the sib was real fast (slow), however, and could finish 30 (or 10) pairs of dots before time was called. Each *S* was asked how many pairs of dots he would *like* to finish (Level of Aspiration, here designated LA) and how many he *thinks* he can finish (Level of Expectation, here designated LE). Half the sib pairs were given the 30 comparison and half the 10 comparison. After stating his LA and LE, *S* was told that this was just a practice game and that he could finish it on completion of all the games. *S* was not allowed to perform on this task so that performance and aspiration on the neutral trials of subsequent tests would not be interfered with. The purpose of Aspiration Task A was to determine the LA and LE under the conditions of *S*'s having no previous knowledge of his own skill.

Subsequent to Aspiration Task A, *S* was shown the first list of Verbal Learning Task B. This task consisted of three lists of pictures, the first six pictures in a list being presented to those *S*s under eight years of age, all eight pictures in a list being presented to *S*s eight years and older. The pictures were India-ink tracings of common objects, chosen from the Stanford-Binet Pictures Vocabulary series (Year II–IV, Form L) and from a series used by Chapman (3). Each list of pictures was presented to *S* for serial learning, the intratrial interval being five sec. and the intertrial interval ten sec. The two response measures obtained for each list on this task were number of trials to learn to one perfect trial (a perfect trial being the correct anticipation of each picture in the list within the five-sec. intratrial interval), and number of errors made during the learning trials. List 1 was presented as a neutral list, while Lists 2 and 3 were preceded by information given to *S* regarding his sib's performance. Half the sib pairs were told before being shown List 2 that the sib had done better than *S* in learning the previous list, and told before List 3 that the sib had done worse. The reverse sequence of sib comparisons was given to the remaining sib pairs. When all three lists were presented under the neutral condition in a pre-experimental test with children aged five to ten years, the means and *SD*s of the three lists were found to be comparable.

---

[4] For a fuller account of the interview method, the testing of its reliability, and of the system of weighting traits to arrive at the trait ratios, the reader is referred to the original doctoral dissertation (6).

The third task, Aspiration Task C, was a pegboard with 11 rows of 60 holes each. For each trial, *S* had to insert as many pegs in a row as he could within 60 sec., using only one hand and picking up one peg at a time. *S* was given two practice trials and nine test trials. A trial consisted of *S*'s stating his LA and LE and then filling in the holes. After three test trials, and preceding Trials 4 through 6, half the sib pairs were told the sib had done better than *S*; preceding Trials 7 through 9 these *S*s were told the sib had done worse than *S*. The reverse sequence of sib comparisons was given to the remaining sib pairs. Standardization of this task under neutral conditions with children aged five to ten years, showed performance on the test to be reliable (the obtained Kuder-Richardson r being .972) and to suffer from only minimal practice effects.

The twenty *S*s used in standardizing the latter two tasks were drawn from pediatric services at a large research hospital (Illinois Research Hospital, part of the University of Illinois medical school). All *S*s, prior to their participation in the present study, were judged by one or more of the medical or nursing staff to be of at least average intelligence and free of gross personality maladjustment. They were either nonbedridden inpatient convalescents (in a ward run as a kind of playground, so that personality effects from hospitalization appeared to *E* minimal for these *S*s) or outpatients coming for their routine pediatric checkup. From their hospital records and their behavior in the experimental situation, these *S*s appeared to be adequately comparable to the experimental *S*s.

Procedure

In the first task, the "sib comparisons" were statements telling *S* how the sib did in comparison to a *norm*, while in the other two tasks the comparisons stated how the sib did in relation to *S*. Although the wording of the original hypotheses covered specifically only the latter situation, for purposes of study the former condition was included, and in subsequent discussion "sib comparisons" will be used to describe both types of comparisons. In summary, the procedure was as follows:

1. *E* played 2 min. with S_2.
2. S_1 was given Aspiration Task A (sib completing 10 (30) pairs of dots).
3. S_2 was given Aspiration Task A (sib completing 10 (30) pairs of dots).
4. S_1 learned List 1 of Verbal Learning Task B (neutral condition).
5. S_2 learned List 1 of Verbal Learning Task B (neutral condition).
6. S_1 was given Aspiration Task C (three neutral trials, three trials sib doing better [worse], and three trials sib doing worse [better]).
7. S_2 was given Aspiration Task C (three neutral trials, three trials sib doing better [worse], and three trials sib doing worse [better]).
8. S_1 learned List 2 of Verbal Learning Task B (sib doing worse [better]).
9. S_2 learned List 2 of Verbal Learning Task B (sib doing worse [better]).
10. S_1 learned List 3 of Verbal Learning Task B (sib doing better [worse]).
11. S_1 did Aspiration Task A if he wished, was praised and given his prize.
12. S_2 learned List of Verbal Learning Task B (sib doing better [worse]).
13. S_2 did Aspiration Task A if he wished, was praised and given his prize.

Comparison conditions were selected for each task by the previous assignment of each sib pair to one of two sequences: (*a*) the 10-Better-Worse sequence and (*b*) the 30-Worse-Better sequence. In the type "*a*" sequence, *S* was told the sib finished 10 dots on Aspiration Task A, the sib did better on Trials 4 through 6 in Aspiration Task C (and worse on Trials 7 through 9) and the sib did worse on List 2 in Verbal Learning Task B (and better on List 3). The type "*b*" sequence is the complete reverse of the type "*a*" sequence. The sequence of comparisons between tasks was thus balanced, i.e., when the comparison was favorable for the sib in the first task (sib finishing 30 dots on Aspiration Task A), the next comparison which *S* received was favorable to *S* (sib doing worse on Trials 4 through 6 on Aspiration Task C), and comparison on the final task was initially favorable again for the sib (sib doing better on List 2 on Verbal Learning Task B). Half each of the Center Competitors, Center Allies, Normal Competitors, and Normal Allies were assigned to the "*a*" condition and the other half to the "*b*" condition. Within each group, the assignment of a given sib pair to the "*a*" or "*b*" condition was random, the only limitation being that no more than half of a given group would fall into one or the other sequence condition.

Response Measures

It will be recalled that in Aspiration Task A no performance measures were obtained and that only one trial was presented to *S*. LA and LE *raw* scores are thus the data used for this task.

For Verbal Learning Task B no aspiration measures were obtained, the two performance measures being number of trials to learn to criterion and number of errors made during the learning trials. The data used for this task were *change* scores from the neutral condition, computed by subtracting the score on the comparison list from the score on the neutral list. A plus score thus indicates improved performance (fewer errors made, or less trials taken, on the comparison list than on the neutral list), a minus score indicating decreased learning efficiency.

In Aspiration Task C, three response measures were obtained: performance (number of pegs inserted within one min.), LA, and LE. As for Verbal Learning Task B, data for this task are *change* scores, a plus score indicating improved performance or raised aspirations. The change score in Task C was computed by subtracting the mean of the three neutral trials from the mean of the three comparison trials.

RESULTS

Aspiration Task A

As seen from Table 2, no significant differences were obtained between Competitors and Allies, although both groups had significantly higher LE when told the sib finished 30 pairs of dots than when told the sib finished 10. Without prior experience in a task, thus, both groups used the sib's performance as a reference point by which to set their own expectations. Their aspirations did not follow this pattern, however. The LA scores tended to be uniformly high, except for the Center Competitors who were told the sib finished 10 dots, the low scores of this group primarily accounting for the significant LA interaction found for Center *S*s. The fact that this significant interaction was not found for the Normal *S*s but only for the Center *S*s, can possibly be explained as a function of the counseling experience Center *S*s have been exposed to.

Verbal Learning Task B

As seen from Table 3, no significant effects were found for this task.

Aspiration Task C

No significant effects were found for the performance measure, although both LA and LE measures yielded significant differences between Allies and Competitors. As seen from

TABLE 2

MEAN LA AND LE SCORES ON ASPIRATION TASK A

(N per cell = 10)

Response Measure	Group	Normal Sample					Center Sample				
		Sib "Fast" (A)		Sib "Slow" (B)		*F*	Sib "Fast" (A)		Sib "Slow" (B)		*F*
		Mean	*SD*	Mean	*SD*		Mean	*SD*	Mean	*SD*	
LA	Competitors	27.9	12.60	31.0	6.63		33.5	6.73	22.9	10.47	
	Allies	30.0	8.94	25.5	10.59		31.0	9.16	33.5	6.87	
	(Between A & B)					.04					1.27
	(Interaction)					1.31					7.77**
LE	Competitors	24.2	9.15	19.3	5.29		31.3	6.80	19.8	9.17	
	Allies	24.8	7.26	17.1	5.10		26.6	6.86	24.3	11.72	
	(Between A & B)					7.46**					5.44*
	(Interaction)					.37					2.42

* $p < .05$.
** $p < .01$.

TABLE 3

VERBAL LEARNING TASK B: MEAN CHANGE SCORES IN NUMBER OF TRIALS TO LEARN TO CRITERION AND IN ERRORS

(N per cell = 20)

Response Measure	Group	Normal Sample				Center Sample			
		Sib "Better"		Sib "Worse"		Sib "Better"		Sib "Worse"	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Trials[a]	Competitors	−.20	2.36	+.05	2.50	+.80	2.29	+1.25	3.00
	Allies	+.65	2.48	+.25	2.74	+.80	3.30	+.25	3.71
Errors[a]	Competitors	−2.10	6.49	−1.00	6.07	−.20	6.24	+1.50	8.07
	Allies	+1.35	6.99	+.05	8.34	+2.00	9.62	−1.70	9.45

[a] None of the obtained *F* ratios were significant.

TABLE 4
MEAN CHANGE SCORES ON ASPIRATION TASK C
(N per cell = 20)

Response Measure	Group	Normal Sample				Center Sample			
		Sib "Better"		Sib "Worse"		Sib "Better"		Sib "Worse"	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Performance	Competitors	+.60	1.51	+.73	1.92	+1.10	1.18	+1.25	1.51
	Allies	+.68	1.49	+.33	1.58	+.87	2.66	+.58	2.46
	(Between Competitor-Ally)	$F = .18$				$F = .96$			
LE[a]	Competitors	+.60	3.85	+1.05	5.16	+.27	3.37	+1.15	3.84
	Allies	+.07	5.72	−3.38	8.44	−.37	9.26	−1.38	9.05
	(Between Competitor-Ally)	$\chi^2 = 2.45$				$\chi^2 = 5.02$*			
LA	Competitors	+3.10	8.75	+2.48	7.63	+3.02	5.88	+.40	7.90
	Allies	−.30	8.05	−2.15	9.59	−3.22	8.14	−2.38	5.26
	(Between Competitor-Ally)	$F = 4.21$*				$F = 8.09$**			

* $p < .05$.
** $p < .01$.
[a] Both for Normal and Center LE, Bartlett's test for homogeneity of variance gave χ^2 significant beyond the .02 level.

Table 4, the Allies lowered and the Competitors raised their LA and LE under both the Better and Worse comparison conditions. Since heterogeneity of variance was found for the LE measure, a chi-square test suggested by Wilson (9) was used for this measure. This yielded a significant difference in LE between Competitors and Allies for only the Center *S*s. Since the results were substantially in the same direction for both Center and Normal *S*s, however, the conclusion is warranted that both LA and LE differentiated between the responses of Competitors and Allies when *S*s were presented with sib comparisons.

When the data of Table 4 are further analyzed, the overall rise in score from the neutral condition seen in the Competitors' LA and LE data turns out to obscure a more complex relationship. As shown in Table 5, for Competitors a significant interaction is obtained between Better-Worse and First-Second presentation on both the LA and LE measures. This interaction describes the fact that the two conditions, of *S* first being told his sib did better and of *S* secondly being told his sib did worse, produce a significantly greater rise in LA and LE than do the two conditions of *S* first being told the sib did worse and of *S* secondly being told his sib did better.[5] This interaction pattern did not appear for the Allies. The presence of this pattern for Competitors suggests that the behavior of the Competitors may be characterized as largely that of rivalry: when Competitors hear that the sib does better, a set to outdo the sib is elicited and maintained even when subsequently the sib does worse, this set to outdo the sib appreciably not being manifested when *S*s first are told the sib does worse.

The meaning of the Allies' drop in LA and LE scores on being presented with the sib comparisons can probably be best understood by inspection of the data of *S*s used to standardize this task. As will be recalled, these *S*s simply took their nine test trials in the manner of the neutral trials of the experimental *S*s, i.e., no comparisons being presented. These standardization *S*s dropped their LA and LE scores from the first three trials to each set of subsequent three trials. In other words, merely performing the task lowers the LA and LE in *S*s of this age, the goal-setting statements becoming closer to the actual performance as a function of increased experience in the task. As seen in Table 6, Allies are comparable to the standardization *S*s in their drop of LA and LE from the first three trials to the subsequent sets of three trials, the introduction of the sib comparisons after the first three trials producing behavior in the Allies that is substantially no different from the behavior of *S*s who were given no comparisons of any kind. This picture is not found for the Competitors, whose change in LA and LE scores from the first three trials differs significantly from the change scores obtained for the standardization *S*s. From this analysis it can be reasonably concluded that Allies, in contrast to the Competitors, were largely unaffected by the sib comparisons, paying heed primarily to their own performance in a manner analogous to *S*s who never were presented with any comparisons. The Allies' lack of apparent concern over the sib's performance suggests that the ally relationship may be weaker than that of competition, stronger cues being necessary to elicit response change in a sib comparison situation. This suggestion, although not specified in the original hypotheses of the present study, is consonant with the Adlerian theory.

---

[5] Since *S*s were randomly assigned to the sequence conditions and since the pattern is similar for Center and Normal *S*s, a selective *S* variable to account for the significant interaction (i.e., that *S*s in the Better First-Worse Second sequence are somehow different from *S*s in the Worse First-Better Second sequence) does not appear to be tenable.

TABLE 5

SEQUENCE EFFECTS FOR COMPETITORS IN ASPIRATION TASK C

(Scores are mean change scores, *N* per cell = 10)

Response Measure	Sample	First Comparison Condition					Second Comparison Condition				
		Sib "Better" First (A)		Sib "Worse" First (B)		*F*	Sib "Better" Second (A)		Sib "Worse" Second (B)		*F*
		Mean	*SD*	Mean	*SD*		Mean	*SD*	Mean	*SD*	
LE[a]	Center	+1.17	3.58	−.20	2.78		−.63	2.88	+2.50	4.25	
	Normal	+1.17	3.87	−1.17	4.54		+.03	3.76	+3.27	4.77	
	(Between A & B)					1.69					7.15**
LA[b]	Center	+3.93	7.56	−1.50	9.53		+2.10	3.22	+2.30	5.15	
	Normal	+6.17	9.69	+.20	4.49		+.03	6.37	+4.77	9.26	
	(Between A & B)					5.55*					4.28*

* $p < .05$.

** $p < .01$.

[a] When Center and Normal scores were combined, *F* for interaction of Better-Worse × First-Second Presentation = 7.59 (for 1,72 *df*, $p < .01$).

[b] When Center and Normal scores were combined, *F* for interaction of Better-Worse × First-Second Presentation = 5.64 ($p < .05$). Although variances were heterogeneous in the *F* test for interaction, a *t* test which compared Better First-Worse Second scores with Better Second-Worse First scores (Center and Normal combined) upheld the significant interaction even though variances were homogeneous for the *t* test.

TABLE 6

MEAN CHANGE SCORES IN LA AND LE ON ASPIRATION TASK C AS A FUNCTION OF TRIAL SEQUENCE[a]

Response Measure	Group	Second Set of Three Trials Minus First Three Trials		Third Set of Three Trials Minus First Three Trials		*F*
		Mean	*SD*	Mean	*SD*	
LA	Competitors	+2.20	8.64	+2.30	6.61	
	Allies	−1.98	6.93	−2.04	8.92	
	Standardization	−2.37	8.24	−3.89	10.25	
	Between Competitors & Standardization					8.39**
LE	Competitors	+.24	3.81	+1.29	4.30	
	Allies	−.84	7.00	−1.69	9.50	
	Standardization	−2.28	11.06	−3.29	11.25	
	Between Competitors & Standardization					5.16*

* $p < .05$. Although variances are heterogeneous, the similarity of trend between LA and LE measures suggests the generality of the Competitor-Standardization difference.

** $p < .01$.

[a] For the experimental *S*s, Better and Worse conditions and Center and Normal *S*s are combined for each cell. *N* per cell thus = 40 for each of the experimental *S*s. *N* for each cell of the Standardization *S*s = 20.

TABLE 7
MEAN SCORES ON FIRST THREE (NEUTRAL) TEST TRIALS IN ASPIRATION TASK C
(N per cell = 20)

Group	Performance		LE		LA	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Competitors						
Normal	24.30	4.88	27.97	9.30	32.75	9.46
Center	25.15	4.08	26.63	5.25	34.54	7.53
Allies						
Normal	25.27	5.69	28.55	8.08	33.42	8.98
Center	23.73	4.03	26.65	4.98	35.00	9.80
Standardization	22.25	4.42	29.12	6.84	34.96	11.58

Since Allies and Competitors differ in their LA and LE once the sib comparisons are presented to *S*, it could be argued that the groups simply differ in their general aspiration behavior, the sib comparisons being of incidental importance. Inspection of Table 7, however, reveals that Competitors and Allies are comparable in LA, LE, and performance on the neutral trials (and are similar to the standardization *S*s), thus ruling out any basic difference between the two groups in general goal-setting behavior.

DISCUSSION

The data failed to confirm the original hypotheses of the study. It will be recalled that a revised formulation, based on the original Adler-Dreikurs conceptualization, was suggested as a result of *post hoc* analysis of the data and task dimensions of the present study. This formulation considered ego-involvement and status-effect as crucial variables in an experimental test of the original hypotheses. In the present investigation, the attempt was made to limit all status manipulation to merely providing *S* with information of the sib's performance, all other sources of ego-involvement and status-effect being minimized.[6] However, on the basis of the two variables of ego-involvement and status-effect, it is possible to distinguish Aspiration Task A from Aspiration Task C: Aspiration Task A only elicits ego-involvement and not status-effect since the information given to *S* compares the sib to a norm and not to *S*'s own performance; in Aspiration Task C, in contrast, the attempt is made to evoke status-effect by telling *S* the sib did better or worse in comparison to *S*'s own performance, the comparisons either deflating or inflating *S*'s status. In addition, Aspiration Task C can be assumed to be more ego-involving, if only because *S*s actually entered the task.

The revised formulation takes into account this distinction between the two tasks. It postulates that under conditions of moderate ego-involvement and status-effect (as in Aspiration Task C) Competitors in their goal-setting behave in a manner analogous to rivalry, while Allies are relatively unaffected by the sib comparisons, paying heed primarily to their own rather than the sib's performance. The latter is taken as indicating a weaker relationship for alliance than competition. The revised formulation further postulates that under low ego-involvement and status-effect (as in Aspiration Task A) the two groups do not differ significantly, although both groups may use information of the sib's performance as a cue for their own strivings if they do not have their own performance available as a cue. The fact that Verbal Learning Task B failed to show significant differences between Allies and Competitors is interpreted as in line with the lack of significant effects found in the performance measure on Aspiration Task C. One explanation for the lack of significant effects in the performance measures is that goal-setting possibly taps motivational variables more directly than do performance measures, so that under conditions of moderate ego-involvement and status-effect (as in Tasks B and C) performance measures fail to distinguish between the behavior of the two groups. Under high degree of ego-involvement and status-effect, however, significant differences in performance should be obtained, as stated in the original hypotheses. These hypotheses, drawn from the real-life experience of the child, which presumably contains a high amount of ego-involvement and status-effect, cannot be said to have been tested in the present study.

---

[6] Although the Adlerian theory of sibling relations describes the relationships of Alliance and Competition as developing as a function of the child's striving for status (4), it was not recognized until the completion of the present study that manipulation of the status dimensions was an essential component to testing the theory. Thus, for experimental rigor, the elicitation of status variables was explicitly limited to merely providing *S* with information about the sib's performance, so that censure and praise, emphasis of success and failure, etc., were purposely minimized so as not to confound the effects of the sib comparisons. Moreover, neutral tasks were chosen both for the purpose of *not* inducing strong success or failure experiences other than those associated with the sib comparisons, and also so that prior experience with a task would not confound the results.

The revised formulation awaits verification from future studies in which ego-involvement and status-effect are systematically manipulated. The roles of various ego-involving instructions and tasks, as well as of status-effect evoked by censure-praise, failure-success defined by means other than the sib comparisons, etc., can then be judged in their importance to the response changes of the two groups. Allies and Competitors who are compared to a stranger introduced for the purpose of the experiment might also profitably be included in further research. According to Adlerian theory, Allies and Competitors would not differ from one another when compared to a stranger as when compared to the sib, for it is postulated that Allies and Competitors do not intrinsically differ in fundamental personality patterns but that they differ in behavior *only* in response to the sib with whom the particular relationship is established.[7] Data from the present study indicate that the two groups were actually closely comparable,[8] suggesting that the change in goal-setting noted in Aspiration Task C can be attributed with fair certainty to the fact that the *sib* was the compared object.

In conclusion, it must be conceded that the present study leaves several unanswered questions. But the fact that siblings, selected solely on the basis of their likeness or oppositeness of personality traits, differ significantly in goal-setting, when the sib's performance is used as the only source of status-effect, suggests the fruitfulness of further experimental investigation of the Adlerian theory of sibling relations.

SUMMARY

The present study set out to test hypotheses derived from the Adlerian theory of sibling relations. Two duplicate experiments were performed, 20 sibling pairs being drawn from Child Guidance Centers and 20 pairs from a "normal" population. Each group contained 10 pairs of Allies and 10 pairs of Competitors, the operational definition of alliance being sameness of traits of the siblings and competition being oppositeness of traits. Diagnosis of the sibling relations was based on interview material obtained from the mother of each pair of sibs. Each *S* was tested singly and was given two aspiration tasks and one verbal learning task. On each task, *S* was given predetermined information about the sib's performance. Although the hypotheses tested by the study were not upheld, Allies and Competitors were found to differ significantly in their goal-setting when the sib's performance was compared to *S*'s own performance. A formulation that revised the original hypotheses of the study was suggested, with the variables of ego-involvement and status-effect being considered crucial to an adequate test of the original hypotheses. Although the present study leaves several unanswered questions, the significant results that were obtained suggest the fruitfulness of further experimental investigation of the Adlerian theory of sibling relations.

---

[7] That is to say that although a given *pair* of Allies will be alike in personality and a given pair of Competitors opposite in personality, a random *group* of Allies will show the same range in personality traits as a random group of Competitors. Thus, out of four Allies, one pair may be striving and dominant and another pair retiring and submissive, while in each of two pairs of Competitors one sib may be striving and dominant and the other sib retiring and submissive. The result will be that two out of the four Allies will be like two out of the four Competitors, and the other two Allies will be like the other two Competitors. In this example, it would be reasonable to expect that the four dominant *S*s will act differently toward a stranger than the four submissive *S*s. Such data would therefore contrast to that of the present study in which the *sib* was the compared object, in that responses to a stranger would not differentiate between the behavior of Allies and Competitors.

[8] Evidence for the comparability of the two groups derives from the following: (*a*) the two groups had comparable scores on the neutral trials of Aspiration Task C; (*b*) the interview data revealed the personalities of the two groups to be quite similar; (*c*) many of the Allies had sibs with whom they had a competitor relationship while several of the Competitors had sibs with whom they had an alliance relationship.

REFERENCES

1. ADLER, A. *The science of living.* New York: Greenberg, 1929.
2. ADLER, A. *The education of children.* London: Allen & Unwin, 1930.
3. CHAPMAN, L. J. An investigation of loss of set in the conceptual performance of schizophrenics. Unpublished doctoral dissertation, Northwestern Univer., 1954.
4. DREIKURS, R. *Fundamentals of Adlerian psychology.* New York: Greenberg, 1950.
5. DREIKURS, R. The psychological interview in medicine. *Amer. J. indiv. Psychol.*, 1952–53, **10,** 99–122.
6. FERGUSON, EVA D. The effect of sibling competition and alliance on level of aspiration, expectation, and performance. Unpublished doctoral dissertation, Northwestern Univer., 1956.
7. HOLT, R. R. Effects of ego-involvement upon levels of aspiration. *Psychiatry*, 1945, **8,** 299–317.
8. IVERSON, M. A., & REUDER, M. E. Ego involvement as an experimental variable. *Psychol. Rep.*, 1956, **2,** 147–181.
9. WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53,** 96–101.

Received April 8, 1957.

EFFECTS OF PUBLIC AND PRIVATE FAILURE ON SELF-EVALUATION[1]

EZRA STOTLAND[2] AND ALVIN ZANDER

Research Center for Group Dynamics, University of Michigan

A PERSON who has failed may as a result lower his evaluation of himself, under certain conditions, yet under other conditions he may not pay much attention to the failure. Among the circumstances that may affect the nature of his reaction are his opinions about how others evaluate his performance, and these opinions are determined by whether the others know about the failure.

It is evident that a person who has failed tends to become sensitive to the opinions of others. This has been shown in several empirical studies (2, 3, 4). It is known, too, that a person's opinion about himself is affected by the evaluations of his performance that he attributes to others (4). If the failer assumes that others think poorly of his performance, his feelings about himself are likely to depend upon how much he attributes validity to the others' judgments about him.

In this study we seek to answer three questions. Two of these are: Will a person's evaluation of himself when he fails in complete privacy be different from his evaluation when he is aware that another knows about the quality of his performance? Will a person's evaluation of himself in the presence of an incompetent judge of his ability be different from his rating when the other person is a competent judge?

In attempting to answer these questions, we take it as given that a person at all times wishes to evaluate himself as highly as it is possible for him to do, within the limits set by the facts as he sees them. When an individual fails on a task and clearly knows that he has done so, he cannot deny his failure to himself without grossly distorting reality. Thus, he is not likely to deny failure in such an instance. But where different degrees of failure are possible and the degree of his failure is uncertain in his own mind, he can deny that he has failed at all, or minimize the degree of the failure.

A person who has failed has no reason to believe that other individuals who are aware of his failure, especially strangers, minimize the extent of that failure. Thus, we assume that a person who has failed believes that others will tend to evaluate his performance as worse than he would like to believe it was.

If a failer, then, conforms to the evaluations he assumes another has of him, he will evaluate his performance low. What will lead him to accept a stranger's negative evaluations? First, it is necessary that the other know about the failure, since unless the other is aware of the poor performance, the failer has few constraints against minimizing his failure to himself. In addition, for him to conform to the other's negative evaluation, the other must generate social pressures on the one who has failed, pressures which are strong enough to overcome the failer's tendency to set his self-evaluation as high as possible. It is assumed that stronger social pressures are exerted by persons who are experts than by those who are nonexperts.

Therefore, it is first hypothesized that, 1(*a*) a person who had failed on a task evaluates the degree of his failure as being greater when his failure is observed by an expert than when it is not observed by an expert; 1(*b*) the observation or nonobservation by a nonexpert, however, does not generate variations in self-evaluations, but 1(*c*) the person evaluates his performance as poorer when his failure is observed by an expert than when it is observed by a nonexpert.

Thus far, the individual has been considered as evaluating only the quality of his performance on a task, as good or poor. He may in addition make evaluations of personal abilities that were involved in the failure. Because his evaluation of his abilities is likely to be affected by the evaluation of his performance, his rating of the abilities will probably be similar to his evaluation of his performance. The three hypotheses described above, therefore, apply equally well to evaluations of abilities involved in the person's performance on the task.

---

[1] The work on this study was financed by a grant-in-aid (M-701 (C)) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

[2] Now at University of Washington.

We conceive abilities as being either closely related or less closely related to the ones under test. Further, we assume that the abilities most closely related to those being employed at the time of his failure will be affected by the level of that evaluation, while those less closely related will be less affected by the evaluation. We raise then a third question: How much of the person's self is included in his evaluations of himself after an experience of failure? A second hypothesis is proposed: Where a person gives a low evaluation of his performance, because of observation of his failure by an expert, the evaluation of closely related abilities will also be low while the evaluation of less closely related abilities will not be as low.

METHOD

These hypotheses were tested in a laboratory experiment. The *S*s were 126 male undergraduate volunteers. As each arrived in the laboratory he was privately given a puzzle which was described as a test of visual motor coordination, with completion of the puzzle as the criterion of success. Before the *S*s worked on it, *E* introduced himself to half of them as an expert on such tests; to the other half as a nonexpert. The *S*s worked in privacy on the task, which was impossible to complete. Then, in half the cases, *E* entered the *S*s' rooms and told them that he observed they had failed (public condition); in the other half, he asked the *S*s, while outside their rooms, to put the test material away without leaving any indication of how well they had done (private condition). All the participants then privately answered a questionnaire in which they evaluated their performances on the task, their visual motor coordination, and their abilities in areas varying in relevance to the task. The separate phases of the experiment are now described.

Experimental Task

The *S*s arrived individually in the laboratory and *E* escorted them immediately to private cubicles where they remained for the rest of the experiment. Each *S* was told he was to work on a puzzle to help him discover things about himself so that he could perform better in the later or "real" parts of the experiment, although he was not told what these "real" parts would be. He was told that working on these tasks would teach him something useful about himself. The purposes of these instructions were to arouse the *S*s' interest in self-evaluation, to motivate them to work on the puzzle, and to interest them in answering the postexperimental questionnaire for their own benefit, not as a communication to a research worker. In this way, it was hoped to obtain more honest and valid answers.

The first task was a puzzle which was presented as "an excellent test of visual-motor coordination." Visual-motor coordination was defined for the *S* as the "ability to imagine how certain shapes and forms fit together and then to take this arrangement and put it down in actuality." Thus, the test was made relevant to certain specified abilities. It was assumed that visual-motor coordination was an aspect of themselves on which these persons did not have fixed evaluations and was, therefore, subject to experimental influence. The puzzle consisted of 12 irregular cardboard pieces with which *S* was to fill the space inside an irregular outline traced on a board. He was allotted ten minutes. The criterion of complete success was said to be full completion of the puzzle. *S* was told that degrees of failure were determined by the discrepancy between the achieved arrangement and the one possible way of completing the puzzle. The number of pieces employed in the design were described as irrelevant to his degree of failure. In this way the objective degree of failure was made ambiguous for the *S*, allowing social factors more influence on his evaluation of his performance.

Before *S* was allowed to work on the puzzle, *E* showed him his second task, the postexperimental questionnaire. This was left in the cubicle for him to fill out after the time allotted for work on the puzzle had been spent. The private nature of his answers was emphasized again by telling him that answers were not of interest to the *E* and by repeating some of the earlier instructions.

Expertness Induction

Before *S* worked on the puzzle, *E* introduced himself in half the cases as Dr. X who had much experience in administering and constructing tests of visual-motor coordination (expert condition). In the other half, *E* introduced himself as a substitute for the regular experimenter and as one who knew little about these tests (nonexpert condition).

Public-Private Induction

The *S* was asked to work privately on the task. The *E* announced that the allotted time had elapsed while standing outside the *S*'s cubicle. In the public condition, *E* entered the cubicle and said impersonally, "I see that you were not able to complete the task. I'm afraid this would have to be called a failure." He then instructed *S* to put the pieces of the puzzle back into the envelope in which he had found them and to go on to the questionnaire. In the private condition, *E* did not enter the cubicle but instead instructed *S* to put the pieces back into the envelope in which he had found them and to go on with the questionnaire. It should be remembered that in both conditions the *S*s knew they had failed to some degree without *E*'s saying so, since they had been told that the criterion of full success was completion of the puzzle.

Thus a four-cell, 2 × 2 design was employed with the following conditions: public-expert, public-nonexpert, private-expert, and private-nonexpert.

Post-experimental Questionnaire

To develop a series of items about self-evaluation that had various degrees of similarity to the abilities involved in the experimental task, a list was obtained (from work as yet unpublished) of the traits that are used most frequently by male undergraduates in describing themselves. This list was ranked a priori by the investigators in respect to their degree of similarity to the abilities needed for the experimental task. To modify this a priori ranking so that it corresponded to the ranking that would be made by an average male undergraduate, a modification of a method used by Coombs (1) was employed, using male undergraduate *S*s other than those in the experiment.

The first item on the questionnaire asked *S* to rate his success on an 11-point scale, if he felt that he had succeeded on the puzzle. The second question was a parallel one for failure, the *S* answering one but not both since all of them had failed. In this way his evaluation of his *performance on the puzzle* was measured. He next made an evaluation of his *visual-motor coordination* (hereafter designated as VMC) and of a series of 13 *abilities* presented in order of decreasing similarity to the abilities required in the experimental task. In the analysis of the results, this 13-item scale was divided into three groups of 4, 4, and 5 items each and the mean evaluation attributed to each group was obtained. These means are referred to hereafter as M1, M2, and M3, the subscripts designating the order of decreasing similarity to the abilities required in the experimental task.

The ratings given to VMC and M1 are assumed to indicate evaluations of abilities that are closely related to the ones used in performing the task, while evaluations of M2 and M3 are assumed to concern areas less related to the task-required abilities. In addition to these 13 items, a final question requested that the *S* evaluate himself as a "complete person."

Finally, they were asked to rate the competence of the *E* on an 11-point scale. This question was inserted in order to determine the extent to which *S*s believed *E* to be an expert or a nonexpert in these two conditions.

RESULTS

To determine whether the *E* was perceived by the *S*s as being either expert or nonexpert, the mean ratings of the competence they attributed to *E* in the post questionnaires were computed. The mean for the expert condition was 7.19, and for the nonexpert condition 5.56, significantly different at the .001 level of probability ($t = 3.62$). It is evident that the *S*s on the whole accepted *E*'s descriptions of his expertness in the separate conditions; but those who were told that the *E* was not an expert attributed more competence to him than might be expected.

TABLE 1

MEAN SELF-EVALUATIONS FOR VARIOUS EXPERIMENTAL CONDITIONS

Item Evaluated	Experimental Conditions				p value of diffs.[*]
	Public Expert $N = 25$ Col. a	Public Nonexpert $N = 39$ Col. b	Private Expert $N = 34$ Col. c	Private Nonexpert $N = 28$ Col. d	
Performance on puzzle	2.58	1.52	2.75	4.20	b–d .001
Visual motor coordination	3.79	4.10	4.78	4.31	a–c .06
Abilities M1	5.07	5.02	5.76	5.37	
Abilities M2	6.88	6.48	6.76	6.92	
Abilities M3	6.11	6.24	6.44	6.22	
Complete person	6.17	6.48	6.71	6.34	

[*] By two-tail t tests.

The mean self-evaluations for their performances on the puzzle and for their separate abilities are contained in Table 1. These data provide no support for hypotheses 1(*a*), 1(*b*), or 1(*c*) in respect to evaluation of performance on the puzzle. In fact, contrary to what was anticipated in hypothesis 1(*b*), the *S*s who were observed by a nonexpert evaluated their performances more poorly than those who were not observed by him, the difference being significant at the .001 level.

The evaluation given to their visual-motor coordination (VMC), is lower in the public-expert condition than in the private-expert condition (p of diff. = .06), as would be expected by hypothesis 1(*a*), and observation or nonobservation by a nonexpert generated no difference in evaluation of VMC as would be expected by hypothesis 1(*b*) These data cannot be accepted as support for hypotheses 1(*a*) nor for 1(*b*), however, since it is apparent that some of these *S*s did not behave as was predicted in evaluating their performances, and the ratings of VMC depend upon a generalization of the evaluation of their performances to their visual-motor coordination as an ability. The evaluations assigned to M1 abilities were approximately the same in all conditions. It is apparent, then, that the data in Table 1 provide no convincing corroboration for hypotheses 1(*a*), 1(*b*), or 1(*c*).

In order to control on the effects of the evaluations given to their performances on the puzzle in determining the evaluations of VMC and M1 abilities, the total sample was divided into two groups at the median of their evaluations of their performances on the puzzle.

TABLE 2

MEAN SELF-EVALUATIONS FOR VARIOUS EXPERIMENTAL CONDITIONS FOR SUBJECTS WITH HIGH AND LOW EVALUATIONS OF PERFORMACE ON PUZZLE

	Experimental Conditions					
Item Evaluated	Public Expert $N = 11$ Col. a	Public Nonexpert $N = 12$ Col. b	Private Expert $N = 18$ Col. c	Private Nonexpert $N = 16$ Col. d	Total Col. e	p value of diffs.*
Part A—Subjects with high evaluation of performance on puzzle						
Visual motor coordination	3.64	5.15	5.17	5.00	4.83	a–c .03 a–b .03
Abilities M1	5.03	5.87	5.97	5.83	5.73	a–b .05 a–c .05
Abilities M2	6.60	7.03	6.87	6.86	6.85	
Abilities M3	6.15	6.68	6.63	6.00	6.26	
Complete person	6.18	7.09	6.72	6.44	6.61	
	$N = 13$	$N = 19$	$N = 9$	$N = 9$		
Part B—Subjects with low evaluation of performance on puzzle						
Visual motor coordination	3.92	3.42	3.89	3.00	3.56	
Abilities M1	5.15	4.49	4.80	4.28	4.68	
Abilities M2	7.13	6.14	6.44	6.44	6.51	a–b .001
Abilities M3	6.52	5.97	6.02	5.96	6.12	
Complete person	6.08	6.11	6.56	5.22	6.20	

Note.—Value of p for differences between comparable cells in Parts A and B of Table 2: VMC, Col. b, .03; VMC, Col. d, .03; VMC, Col. e, .001; M1, Col. b, .001; M1, Col. c, .05; M1, Col. d, .03; M1, Col. e, .001.

[a] By two-tail t tests.

Then, the means of the evaluations given to their separate abilities were calculated. These are reported in Table 2.

It can be seen there that those *S*s who minimized their failures and gave better evaluations to their performances (Part A), tended in the large majority of cases to give higher evaluations to their abilities than did *S*s (Part B) who rated their performances as poorer. There is a strong tendency, it is clear, for persons who think better of their performances to also think better of their abilities. This tendency is evident in every experimental condition, except the public-expert one.

In the public-expert condition, the *S*s who had rated their performances high evaluated their VMC and M1 abilities as low as did the *S*s who had rated their performances low. We return later to a discussion of this discrepancy in the evaluations of performances and abilities in the public-expert condition.

The low evaluations of the VMC and M1 abilities by *S*s in the public-expert condition, but not in the other conditions, provide data relevant to the first hypothesis.

The (*a*) part of Hypothesis 1 asserts that persons make lower evaluations of themselves when their failures are observed by an expert than when not observed. It is clear that VMC and M1 abilities are rated significantly lower by *S*s who were in the public-expert conditions than by those who were in the private-expert condition. This is true, however, only for those who evaluated highly their performances on the puzzle.

The (*b*) part of Hypothesis 1 states that where the *E* is a nonexpert, persons do not evaluate themselves differently when their failure is observed as compared to when it is not observed. The hypothesis is supported both by those who rated their performance high, and by those who rated their performance low.

The (*c*) part of Hypothesis 1 says that observation of the *S*s' failures by an expert generates lower self-evaluations than does observation by a nonexpert. It is evident that this hypothesis is supported by the responses of those *S*s who gave themselves high evaluations in their performance on the puzzle, but is not supported by the *S*s who had low evaluations of their performance on the puzzle.

It is clear, then, that two of the three parts of the first hypothesis are supported by ratings given to VMC and M1 abilities only where the *S*s felt that, or stated that, they had performed well on the puzzle.

Among the *S*s who made low evaluations of their performances (Part B of Table 2) there are no significant differences in the evaluations of VMC and M1 abilities when the means of the four experimental conditions are compared. Perhaps these *S*s evaluated their performances on the puzzle, and their VMC and M1 abilities, so low that the pressure on them, stemming from the *E* to lower their evaluations of themselves could cause no further drop in self-evaluations regardless of the strength of these pressures. Where the *S*s had made high evaluations of their performances on the puzzle, the height of these evaluations apparently provided "room" for the social pressures effectively to lower evaluations of closely related abilities.

Hypothesis 2 states that where a person gives a low evaluation of his performance because his failure has been observed by an expert, the evaluation of closely related abilities will also be low, while the evaluations of less closely related abilities will not be as low. The results in Table 1 are inconclusive in respect to this hypothesis, although they are in the predicted direction.

In Part A of Table 2, as we have seen, the abilities which are most closely related to performance on the puzzle (VMC and M1) were evaluated significantly lower where the *E* was an expert and observed the *S*s' failures and where the *S*s had placed a high evaluation upon their performances on the puzzle. It is evident that abilities M2, M3, and the "complete person" were not, however, rated lower where *E* was an expert and observed the *S*s' failures. These results support Hypothesis 2. In Part B of Table 2, *S*s also evaluated their M2 abilities lower, for some unexplainable reason, when their failures had been observed by a nonexpert rather than an expert.

It is also relevant to note that participants who gave low evaluations to their performances on the puzzle, compared to those who gave high evaluation to their performances, attributed significantly lower evaluations to their VMC and M1 abilities when the responses of *S*s from all conditions were combined (Col. e in Table 2). There were no significant differences, however, in the ratings given to M2, M3, or the "complete person" (all conditions combined) when the responses of persons with high evaluations of their performance were compared with those who had low evaluations of their performances. These results appear again to indicate that low evaluation of one's original performance affect evaluations of the most closely related abilities more than they affect evaluations of those abilities which are less closely related.

DISCUSSION

The results suggest a modification of the assumptions and hypotheses, but a modification that is consistent with the rest of the theorizing. In the statement of the theory at the outset of this paper, it had been assumed that individuals would tend to evaluate themselves as highly as possible within the limits of reality, but that pressures toward more objective evaluations from other persons would cause the *S*s to depress their evaluations of their performances and abilities. The data suggest, however, that some persons who fail on a task make such low evaluations of their performances on that task, and of their abilities related to their performances, that social pressures could not lead to any further depression in their self-evaluations. Thus, it was found that *S*s who made low self-evaluations of their performances evaluated their closely related abilities uniformly low, irrespective of whether they were being observed by the *E* and whether he was an expert.

On the other hand, the high evaluations that some *S*s made of their performances can be taken as a sign that they were motivated to make high evaluations of themselves, more than were those who made low evaluations of their performances. For the highly motivated *S*s, variations in the degree of publicness of the *S*s' failures and in the expertness of the *E* had the predicted consequences on their evaluations of abilities. When they were observed by an expert, they conformed to his presumably stronger influences and to his evident low evaluations, and thus evaluated their abilities as poorly as did the *S*s with low evaluations of their performances. In line with the original hypotheses, those *S*s with high evaluations of their performances rated lower abilities that were more closely related to the performances on the puzzle when observed by an expert than when not observed by him, or when observed by a nonexpert. Also, in keeping with the hypotheses, their evaluations of abilities were not different when they were observed by a nonexpert than when not observed by him. Where the assumption that a person will try to keep his self-evaluation as high as possible is shown by the data to be present and operating, then, the experimental hypotheses were supported in regard to the *S*s' evaluations of their relevant abilities. Where this assumption was not fulfilled, the hypotheses were not supported. Instead of assuming, as was done in the opening paragraphs of this paper, that there is a tendency in all persons at all times to evaluate themselves as highly as possible, we must assume that this tendency is present in some persons. Where it is present, the hypotheses proposed for this study are best fulfilled.

This modification of the assumptions and hypotheses raises the question as to why some *S*s attributed high evaluations to their performance on the puzzle, while others did not. In particular, the question is raised in reference to the *S*s in the public-expert condition in Table 2. No conclusive answer to this question is available, although it is likely that some of the *S*s felt more need to look good in the eyes of the *E* even when they made their ratings privately. It is possible that the high self-ratings in the public expert condition may have been due to resistance to the stronger pressures exerted by the *E* in that condition. A complete answer probably requires some analysis of the motivations of the *S*s when rating themselves.

These results point to the need for a theoretical and experimental refinement of the meaning of "expertness" for interpersonal relationships. It is not clear from the present experiment whether it is the prestige of the expert, the power of the expert, or the objectivity he assumes, which generates the obtained differences in the evaluations of the *S*s in the separate conditions. A crucial test would be to obtain self-evaluations from *S*s who were observed by a person whose prestige was varied independently of variations in his apparent ability to be competent in judging the *S*s' performances.

The results support the second hypothesis of the study, namely, that the differences among the conditions are limited to evaluations of abilities most closely related to the abilities required for performance of the task. Only one significant difference was found among the conditions in the evaluations of other areas of the self, in all probability a result without psychological significance.

SUMMARY

In this study we have examined how variations in the strength of social pressures, acting upon a person in the direction of lowering his self-evaluation, affect the evaluations the individual assigns to the quality of his performmance, and to his more or less closely relevant abilities, after he has failed on a task. Social pressures were assumed to be stronger on the person when his performance was viewed by an expert than when it was not observed, or when it was observed by a nonexpert. Social pressures were assumed to restrain a person from assigning a high evaluation to the quality of his behavior or abilities.

The subjects were individually and privately each asked to solve an impossible puzzle which was described as a test of visual-motor coordination. In half of the cases their level of performance was observed by the experimenter; in the other half it was not observed. Within each of these conditions the experimenter introduced himself to half the subjects as an expert and to the other half as a nonexpert. After the subjects had failed, they rated their performances and abilities. The abilities had varying degrees of similarity to those under test in the puzzle.

It was found that:

1. *S*s who made high evaluations of their performances on the puzzle, evaluated abilities that were involved in their performances lower when they had been observed by an expert than when they had not been observed by an expert, or when they had been observed by a nonexpert.

2. *S*s who made low evaluations of their performances did not evaluate these abilities differently under different experimental conditions.

3. No consistent consequences from the publicness or privacy of failure, nor from the expertness of the experimenter were found when the results from the entire population of subjects was examined in respect to the

evaluations of the quality of their performances and their abilities.

4. The differences in the effects of the experimental conditions were limited to those abilities which were most closely related to the abilities required by the task; evaluations of other parts of the self were less affected.

It was concluded that strong social pressures acting on a person toward lower evaluations cause more lowering in his evaluation of his abilities when the person is strongly motivated to keep his evaluations high, than when he has little motivation to keep these evaluations high.

REFERENCES

1. Coombs, C. H. A method for the study of interstimulus similarity. *Psychometrika*, 1954, **19**, 183–194.
2. Jackson, J., & Saltzstein, H. D., *Group membership and conformity processes.* Res. Center for Group Dynamics, Ann Arbor, 1956.
3. Rassmussen, G., & Zander, A. Group membership and self-evaluation. *Hum. Relat.*, 1954, **7**, 239–251.
4. Stotland, E., Thorley, S., Thomas, E., Cohen, A. R., & Zander, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55–63.

Received April 23, 1957.

DISTANCE AS A DETERMINANT OF INFLUENCE IN A TWO-PERSON SERIAL INTERACTION SITUATION

SEYMOUR FISHER AND ARDIE LUBIN[1]

Walter Reed Army Institute of Research

THIS paper is concerned with two major aspects of a two-person social influence situation. The first deals with the systematic exploration of the effects of "distance" upon measures of influence, where distance refers to the size of the discrepancy between the subject's (*S*'s) pre-influence judgment and the subsequent judgment originating from an influence source (6). The concept of distance corresponds closely to "opinion deviance" as used by Festinger (2), Schachter (20), Goldberg (8), and others; in Asch's work (e.g., 1), he speaks of "size of discrepancy"; French (7) employs the same phrase in his model. These writers, along with Newcomb (18), have all suggested in one form or another that distance is an important variable in determining how much *S*'s judgment can be altered in a social interaction. The precise relationship, however, seems to be in doubt. Do large distances produce more "influence" than smaller distances? One difficulty in attempting to answer such a question lies in the ambiguity of the term "influence." As noted in a previous paper (6), influence can be measured in at least two ways: (*a*) how much *S moved* from his original judgment, and (*b*) how much *S conformed* to the source's judgment (i.e., *S*'s movement relative to the distance). So long as distance is held constant, measures of movement and conformity are in a one-to-one relationship; hence, either can be used satisfactorily. When *S*s are exposed to different distances, however, the two measures are no longer directly comparable and yield different results. An *S* who moves 100 units away from his pre-influence judgment under a distance pressure of 200 is behaving quite differently from another *S* who moves 100 units under a distance pressure of 1000; while both *S*s have *moved* equal amounts, the former *S* has *conformed* .50 (i.e., has moved half-way toward the influence source's judgment) as contrasted with the latter *S* who has conformed only .10. One of the purposes of this study was to investigate parametrically the way in which movement and conformity are related to the independent variable of distance.

The second problem under investigation concerns behavior in a *serial interaction situation*, where *S* is exposed to a number of influence attempts on successive, discrete stimuli, each common to the same broad dimension. Whereas most experimental studies seem to focus on the effects of one or more influence trials treated as a single influence attempt (measuring the difference between pre- and post-influence judgments), it is more usual in real life that *S*s are exposed to a set of influence attempts from one source, sometimes on the same issue, sometimes on different issues. In the present study, the concept of "serial interaction" was experimentally simplified by studying *S*s' reactions to two consecutive stimuli. Three questions were being asked here: (*a*) What are the effects of repeated influence attempts on the *same* stimulus? (*b*) What carry-over effects does a series of influence attempts on one stimulus have upon the perception of a second, similar stimulus? (*c*) What effects does a series of influence attempts on one stimulus have upon subsequent influence attempts on a second, similar stimulus?

METHOD

The data have been obtained from two separate studies, the second essentially a replication of the first. The *S*s were army enlisted men from Fort Meade, Maryland. Both studies required *S*s to judge the number of paratroopers shown in briefly exposed photographs. In the original experiment, 96 *S*s were used in a 3 x 2 x 2 factorial design. The independent variables consisted of three different distances (40, 320, and 5100), two experimenters (*E*s), and two sets of instructions (one emphasizing accuracy, the other emphasizing agreement with partner). A pilot study showed that a significant experimenter-distance interaction would result unless instructions were rigidly standardized; in the study, instructions, *E*s, and interactions were analyzed—none was significant. The second experiment used a total of 192 *S*s, divided into six distance groups (0, 40, 160, 320, 1300, 5100), 32 *S*s per group; two different *E*s ran the study, with standardized instructions emphasizing accuracy.

Two *S*s at a time were brought into the laboratory

---

[1] Grateful acknowledgment is made to the invaluable technical assistance of Esther Adams, Nicholas Covatta, Peter Kyriazis, Major E. E. Admire, and Hilda Wexler.

and seated side by side in front of *E*, who read the preliminary instructions to them. In brief, these instructions said that *E* was interested in seeing how accurately people could judge the number of troops shown in photos for short periods of time; further, whether people could be more accurate when they were working together and able to talk things over than when they were by themselves and could not discuss their judgments. Each *S* was given a pad of colored paper to write down his judgments, the color of the pad serving as identification of *S*.

"Warm-up" Period

E began by exposing a photo for 10 seconds. After the photo was removed, the two *S*s wrote their independent estimates on their pads. The estimates were announced to the *S*s, and (if *S*s disagreed) they were told to "talk it over between yourselves and try to be as accurate as possible" (accuracy instructions). After a few minutes of discussion, each wrote down his second estimate, after which *E* announced the true number, and *S*s were permitted to examine the photo and verify *E*'s statement. This warm-up period was extended over five photos and served a number of important functions: (*a*) *S*s became accustomed to estimating the number of paratroopers, and also became familiar with the procedure of making more than one estimate to the same photo. (*b*) The credibility of the subsequent experimental phase was enhanced. Each time *S* wrote down an estimate during the warm-up, the papers were taken by *E* to an assistant a few feet away (and partially out of sight) who recorded and audibly stamped the estimates by machine. Thus, in the experimental phase which followed, *S*s accepted the routine of having *E* take their estimates to the assistant *before* they were exchanged (permitting the assistant to intercept the original estimates and exchange them for bogus messages).

The *S*s were then told that for the remaining time they were to work alone, with no talking between them; all they could do was exchange their estimates. They were placed in separate, adjacent cubicles where they could not see each other. Each *S* retained the same colored pad he had used in the "warm-up."

Experimental Period

In the account of the procedure which follows, the major measures which were analyzed are indicated in parentheses immediately following the operations to which they refer.

E exposed Card A, a photo containing 165 paratroopers, for a 10-second interval, and *S*s wrote down their initial estimates (A_1). *E* picked up the estimates to be stamped and recorded, during which time the bogus estimates were written by the assistant on appropriately colored paper. Hence, when the estimates were "exchanged," each *S* was led to believe that he was receiving his partner's estimate, whereas he actually received an estimate which was a fixed distance *greater* than his initial estimate (except, of course, that *S*s in the zero distance group received bogus estimates identical with their A_1). *S*s then made their second estimate (A_2), the estimates were again "exchanged," and a total of five estimates was made by each *S* to Photo A. *The bogus estimates always remained the same* for any *S* throughout the five interchanges, regardless of whether or not *S* changed his estimate at any time. After the fifth estimate, Photo A was exposed again for 10 seconds, and *S*s were asked to make a "final" estimate (A_u).[2] A_u and A_1 were the only estimates made immediately after seeing Photo A, without an intervening bogus estimate.

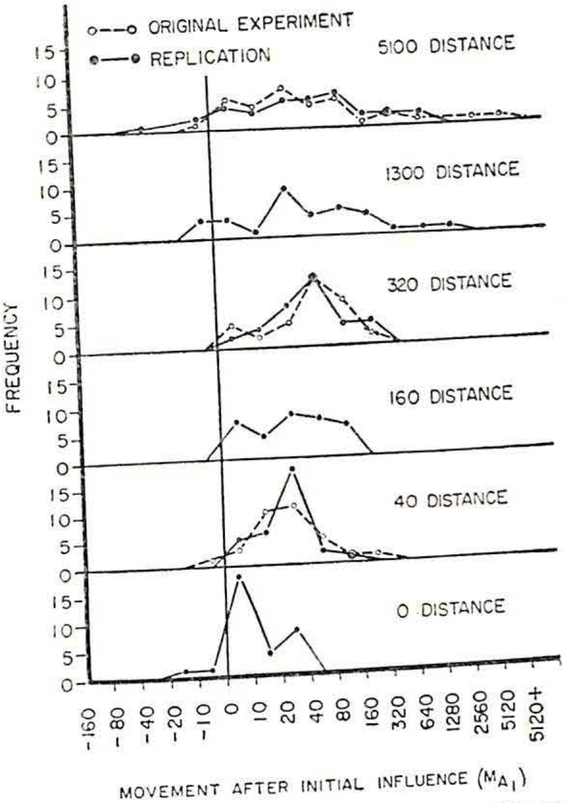


FIG. 1. FREQUENCY DISTRIBUTIONS OF INITIAL MOVEMENT SCORES ON CARD A AS A FUNCTION OF DISTANCE (Note abscissa is scaled geometrically)

Card B, a photo containing approximately 300 paratroopers, was then shown for 10 seconds, and *S*s made their initial estimates (B_1) to this photo. The messages were "exchanged," but this time *all S*s, regardless of the distance received on Photo A, *received a bogus estimate 320 greater than* B_1. The *S*s made a second estimate (B_2), after which Photo B was again exposed for 10 seconds, and *S*s made a "final" estimate (B_u). An additional measure, taken only in the replication study, consisted of giving *S*, at the end of the experiment, Photos A and B for a period of three minutes during which time he was asked to count the actual number of paratroopers in the two photos (A_v and B_v, respectively).

---

[2] In the replication, half of the *S*s were not given the A_u exposure at this point; instead, it was delayed until after *S* made his B_2 estimate (vide infra). For all of the measures analyzed, this variation in treatment produced no significant effects, hence the A_u estimates have been pooled.

RESULTS

The First Stimulus (*Card A*)

Intratrial effects. In the analysis of these data, two separate measures of influence were utilized. *Movement* is the amount of change from *S*'s initial estimate (A_1) to a subsequent estimate for the same stimulus; thus M_{A_1} indicates *S*'s change on Card A from A_1 to A_2; M_{A_4} indicates the change from A_1 to A_5 (last estimate). *Conformity* is *S*'s change proportional to the distance between him and his partner; thus, for an *S* who was exposed to a distance of 40 on Card A, C_{A_1} was computed by dividing M_{A_1} by 40; C_{A_4} was computed by dividing M_{A_4} by 40.

Fig. 1 shows the frequency distributions of M_{A_1} for the six distances used (note that the abscissa is scaled geometrically for convenience). At large distances, distributions take on a marked positive skew. Similar skewed distributions obtain for virtually all variables measured after exposure to different distances. Because of this skew, the data were analyzed using the median and other nonparametric statistics.

In Fig. 2, the basic relations between influence and distance are presented. Influence on Photo A is measured by movement and conformity after the first influence attempt in Figs. 2a and 2b, after the last influence attempt in Figs. 2c and 2d. The following facts may be gleaned from the curves:

1. The results are quite stable, as evidenced by the similarity of the data obtained from the two experiments (all medians are based upon an *n* of 32).

2. The median amount of movement increases with distance—at least up to a certain point, after which a decreasing trend is suggested. One might expect, a priori, that for a particular individual, movement will increase with distance until a point of subjective "incredulity" or "implausibility" is reached, above which the influence source will be disre-

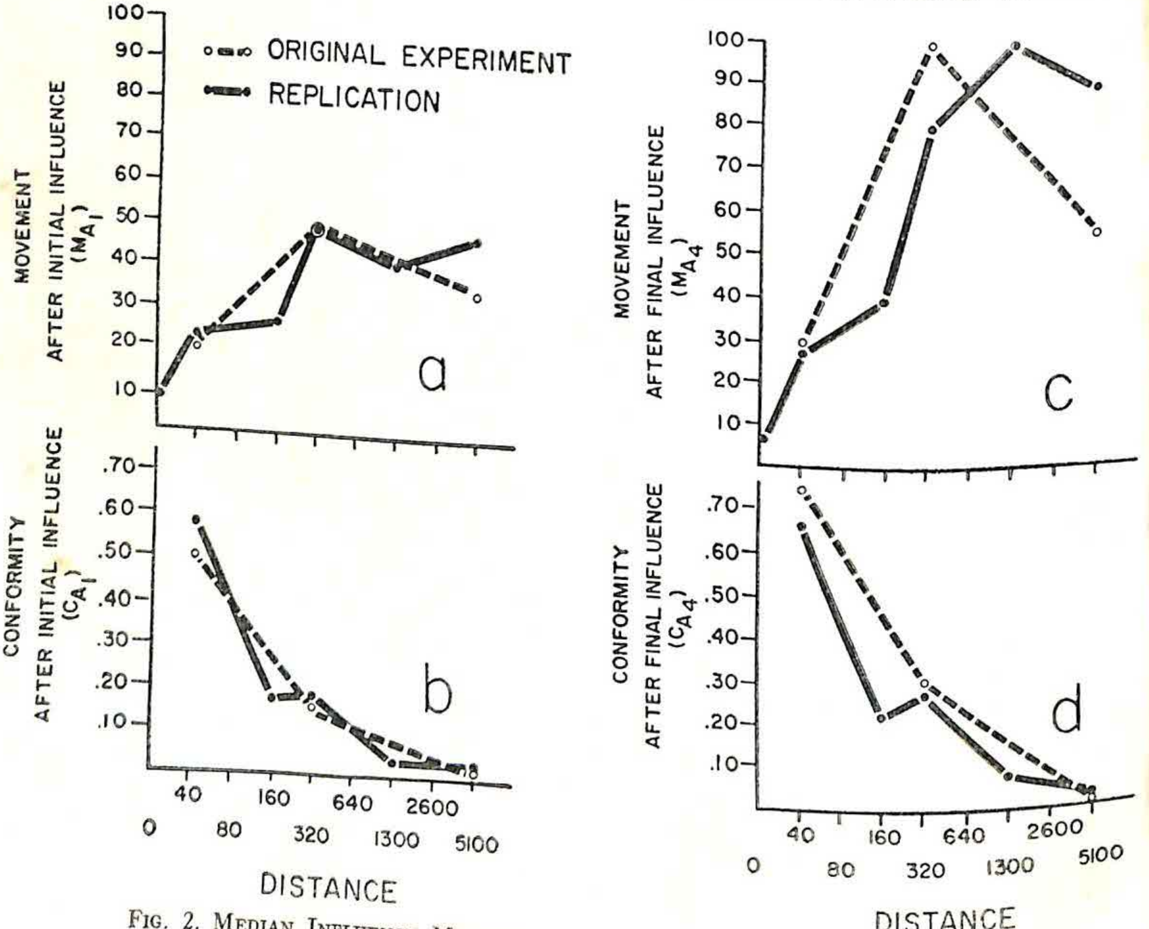


FIG. 2. MEDIAN INFLUENCE MEASURES ON CARD A AS A FUNCTION OF DISTANCE

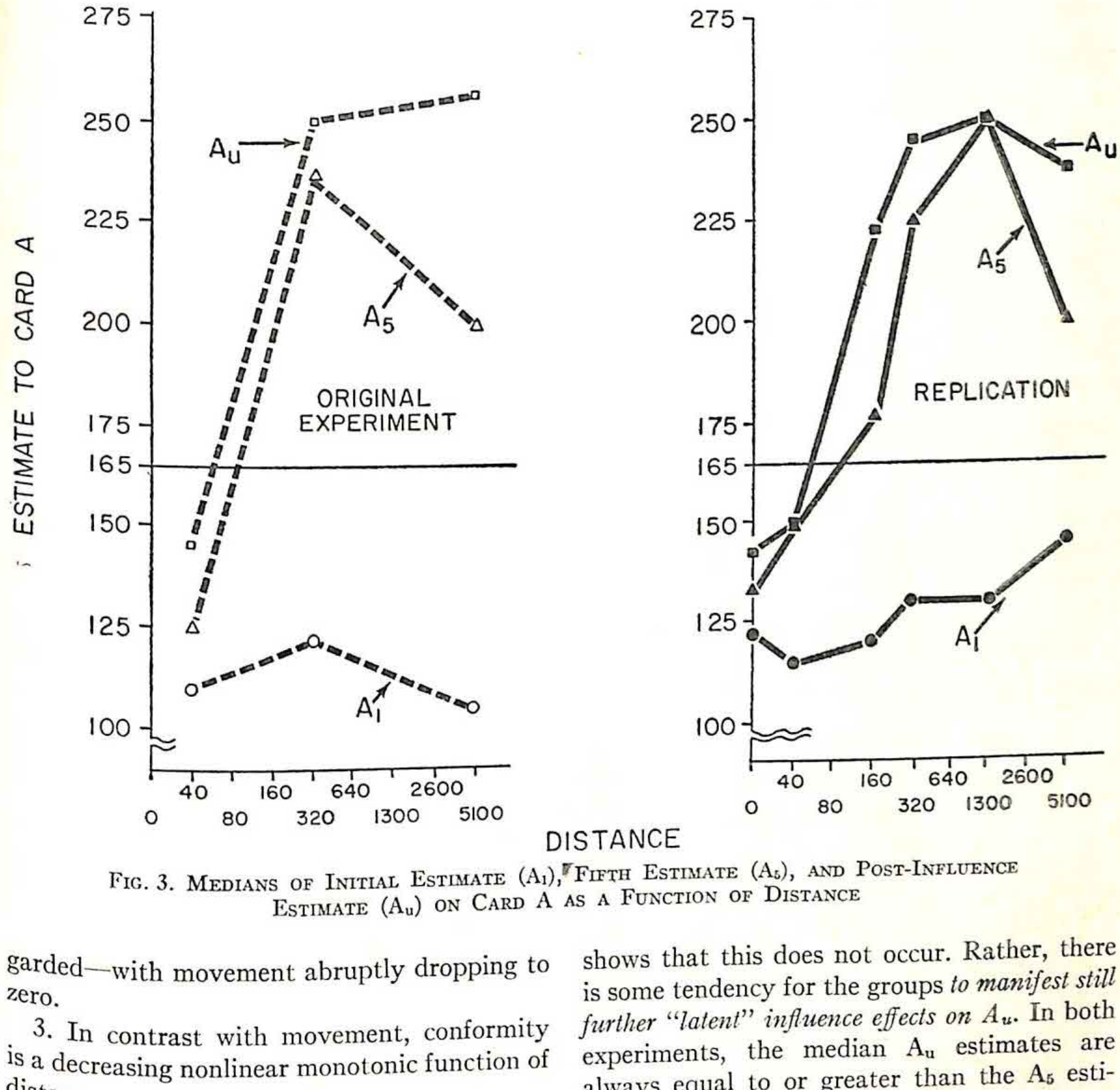


Fig. 3. Medians of Initial Estimate (A_1), Fifth Estimate (A_5), and Post-Influence Estimate (A_u) on Card A as a Function of Distance

garded—with movement abruptly dropping to zero.

3. In contrast with movement, conformity is a decreasing nonlinear monotonic function of distance.

4. With repeated influence attempts on the same stimulus, the amount of influence (both movement and conformity) is increased. This is particularly true of movement scores, seen at all distances; C_{A_1} and C_{A_4}, however, are essentially the same at the larger distances. Maximum influence generally occurs on the first influence attempt (cf. 8, 9).

Persistence of influence. It will be recalled that after the five estimates on Photo A were "exchanged," *S* was again shown the same photo for 10 seconds and asked to give his "final" estimate (A_u). It was expected that, for those *S*s who had been influenced from A_1 to A_5, an additional exposure to "reality" would permit them to correct themselves and give more accurate estimates. Fig. 3, however, shows that this does not occur. Rather, there is some tendency for the groups *to manifest still further "latent" influence effects on A_u.* In both experiments, the median A_u estimates are always equal to or greater than the A_5 estimates. Approximately 40 per cent of all *S*s moved from A_5 toward the bogus estimate (and *away* from "reality") on A_u, while another 10 per cent (whose A_5 estimates were all below 165) also moved toward the bogus estimate, coming closer to the true number; 25 per cent of the *S*s did not change their A_5 judgments at all, and the remaining 25 per cent moved away from the bogus estimate back toward "reality."

While it appears, then, that influence persists when *S* is re-exposed to the same ambiguous situation (without any direct influence attempt being made), it should not be inferred that these effects represent an enduring change. When *S* is permitted a relatively extensive reality contact with the stimulus (i.e., giving him three minutes to count at the conclusion of the study), the A_v estimates do tend to drop back toward the true number, regardless of previous distance-experience: *S*s seem capable of shaking off the influence effects to a considerable degree when the ambiguity of the stimulus is removed, although the over-all median A_v (184) still is somewhat higher than the true number (165). This apparent overestimate is difficult to interpret clearly, since an adequate control group is lacking (*S*s in the zero-distance group were exposed to a distance of 320 on Photo B *prior to* their A_v estimates, and therefore do not provide an appropriate comparison); before one could validly infer that the elevation of the obtained A_v curve indicates some persistence of influence, it would be necessary to show that *S*s in the absence of any previous influence attempts are able to estimate accurately (or tend to underestimate) Photo A during a comparable three-minute counting session.

The Second Stimulus (Card B)

The intertrial phenomenon. In an earlier study of continuous social influence interactions (6), it was found that under certain conditions of immediate self-committal (i.e., where *S* gives an explicit judgment just prior to an influence attempt), a considerable amount of influence may be reflected *between* rather than within interactions (trials). When *S* has been under continuous pressure, applied in some consistent pattern, it is not uncommon to observe *S* appearing to anticipate the influence source on subsequent interactions. Marked changes are seen occurring in *S*'s committal responses themselves, prior to any overt attempt to alter his judgment on any particular stimulus.

It was expected that the intratrial pressures exerted on *S* during the five influence attempts on Photo A would be reflected by a systematic increase in the initial estimate (committal response) to Photo B (B_1). From Fig. 4 it is clear that this indeed occurs. B_1 is significantly related to the distance *S* was exposed to on Photo A: the larger the distance on Photo A, the larger the median B_1. Groups exposed to large distances on a stimulus may be highly resistant to intratrial influence on that particular stimulus, but when exposed to a new stimulus (prior to any direct influence attempt), the previous pressures become patently evident. A number of *S*s in the 5100-distance group showed little intratrial movement on Photo A, yet gave initial estimates of 4500 or higher to Photo B (containing 300 paratroopers).

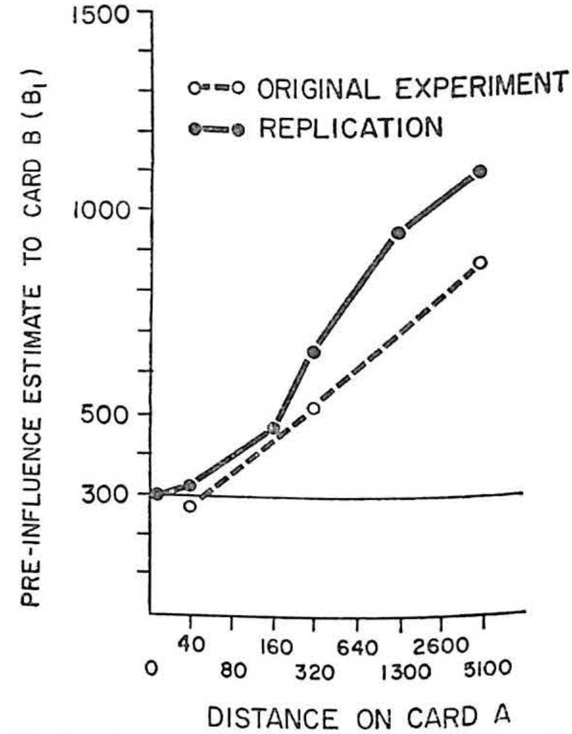


FIG. 4. MEDIAN INITIAL ESTIMATE (B_1) ON CARD B AS A FUNCTION OF DISTANCE ON CARD A (INTERTRIAL EFFECT)

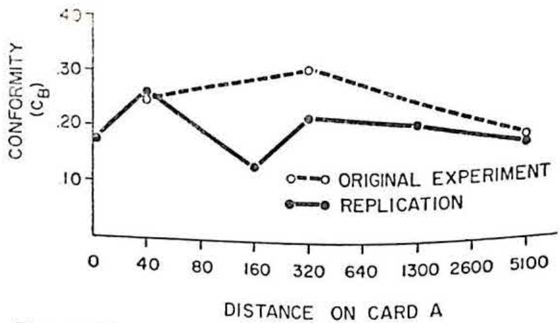


FIG. 5. MEDIAN INTRATRIAL CONFORMITY (C_B) ON CARD B AS A FUNCTION OF DISTANCE ON CARD A

Intratrial effects. What happens when *S*s who have been exposed to different distances on one stimulus are subsequently exposed to a uniform distance (320) on a second stimulus? It has already been seen that their initial estimate (B_1) of this second stimulus is significantly related to their previous distance-experience (intertrial effect). But after they have

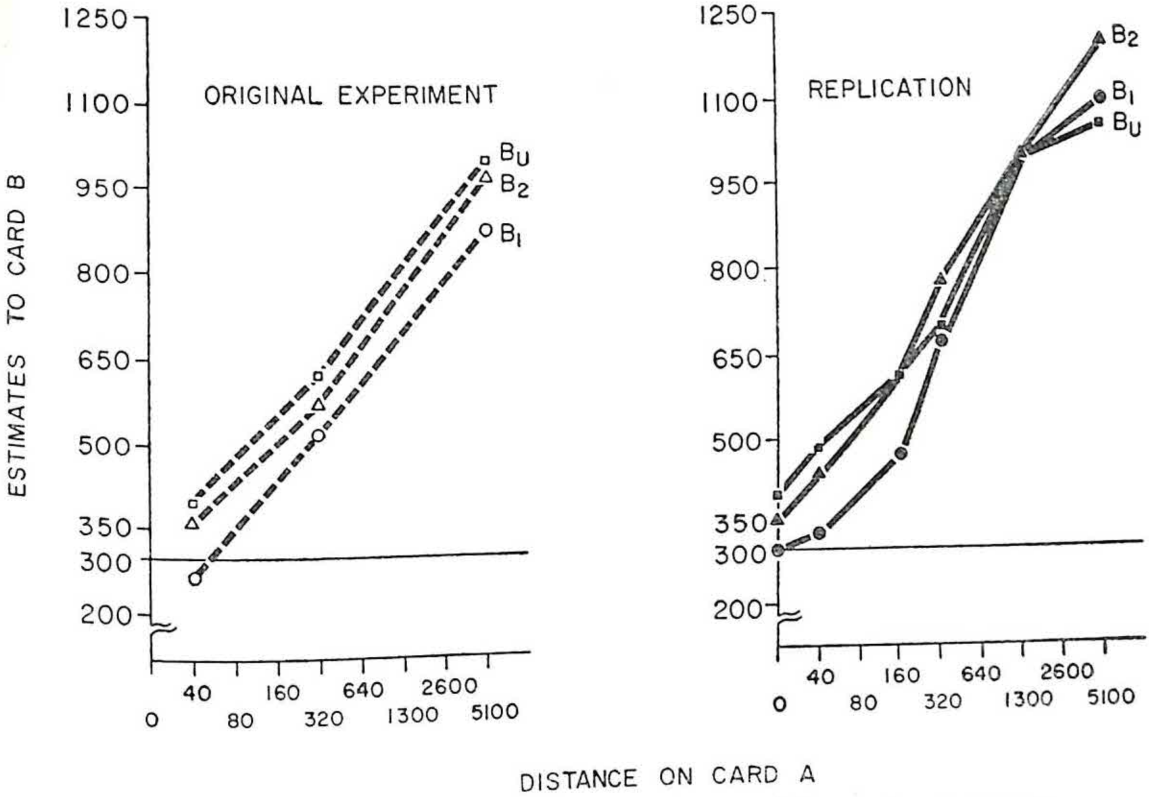


Fig. 6. Medians of Initial Estimate (B_1), Second Estimate (B_2), and Post-Influence Estimate (B_u) on Card B as a function of Distance on Card A

committed themselves to Photo B, how easily can their judgments be further altered? Will *S*s who were exposed to a distance of 5100 on Card A behave differently from *S*s who were exposed to zero distance? Fig. 5 indicates that there is no systematic relationship between C_B, the amount of conformity on Photo B, and distance on Photo A. Despite the fact that the 5100-distance group had already demonstrated intertrial influence on Photo B, this group appears to change neither more nor less than any other group—including the zero-distance *S*s who show essentially no intertrial effects.

The rank-order correlations of C_B with several other variables (e.g., C_{A1}, A_1, A_5, B_1) were computed within each distance-group, but no significant relations were obtained. Under the conditions of the present study, then, C_B cannot be demonstrated to be related to any of the temporally previous variables measured, including distance on Photo A.

Persistence of influence. When *S*s were given a ten-second re-exposure to Photo B (B_u), Fig. 6 reveals a marked tendency toward maintenance of the previous influence effects, similar to the results on Photo A. The B_u curve closely follows the B_2 curve at all distances: *S*s do *not* show any strong indication of dropping back to "reality" even at the extreme distances. However, the "latent" influence effect on B_u is not as clear as for A_u.

When given three minutes to count Photo B at the conclusion of the study, *S*s' estimates (B_v) are now independent of the distance on Photo A. All groups have similar median B_v judgments (approximately 500), but all medians are considerably above the true number of paratroopers (300). As noted in the case of A_v, interpretation is equivocal in the absence of proper controls, but there is certainly the suggestion that the "counting" behavior of these *S*s may be modified by previously applied pressures. Since it was virtually impossible to make an accurate count of the 300 paratroopers on Photo B (quite different from Photo A, which contained only about half the number), the B_v measurement becomes in effect a continued influence situation—differing from B_u (seen for only 10 seconds) only in the degree of stimulus ambiguity. A few *S*s in the

1300 and 5100 distance groups who had given high B_1 and B_u estimates were observed to go through the complete motions of "counting," and then give B_v judgments of 3500 or 4000!

DISCUSSION

The Effects of Distance

One of the major questions in this research concerned the relationship between measures of social influence and distance. The results clearly indicate that while movement tends to increase (at least up to a certain point) with distance, conformity is a monotonic nonlinear decreasing function. It would seem, then, that distance is a significant determinant of social influence, and hence must be considered in any attempt to formulate theoretical models of influence—particularly since a very recent paper by Hovland and Pritzker (11) reports similar relations using attitudinal stimuli low in ego-involvement. These results do not appear to support Goldberg's earlier finding (8) that conformity is independent of distance. The present results are also pertinent to the Osgood and Tannenbaum model (19), French's theory (7), and Mausner's (17) recent discussion.

Intertrial Effects and Persistence of Influence

A striking phenomenon consistently observed in recent studies of social influence in continuous interactions (6) concerns the large amounts of influence manifested *between* interactions (i.e., in this instance, stimuli A and B). In the present experiment, *S*s on Photo A not only tended to remain influenced when the *same* stimulus was presented a second time (A_u), but also showed a carry-over effect when a new stimulus (Photo B) was presented: *S*s demonstrated gross over-estimation tendencies positively related to the distance-pressure received on the first stimulus. These intertrial effects, occurring prior to any overt, direct influence attempts on the second stimulus, often seem to rest in "anticipatory" attitudes in *S*. Just as expectations play an important role in hypnotic behavior (3, 4), similar factors seem to be operative in nonhypnotic continuous influence situations. To the extent that *S* can perceive a more or less consistent bias in the source's judgments, his frame of reference (22) is altered, and he subsequently tends to manifest indirect influence effects in giving his initial, pre-influence (committal) judgments to new stimuli. Some *S*s may even be quite resistant to intratrial pressures on the first stimulus, and then show a large intertrial effect. (This phenomenon may be a generalization of the "sleeper effect," noted in a number of attitude-influence studies [e.g., 10, 13]: the main difference seems to be that whereas the "sleeper effect" refers to increased influence observed in subsequent retesting on the *same* stimulus, the intertrial effect refers to increased influence seen during testing on a *new* but related stimulus.)

When the stimuli all belong to a quantitative dimension, as in the present study, it is probable that some psychophysical "anchoring" effects from *S*'s previous judgments partly contribute to the intertrial change—i.e., if an *S* was actually influenced on Photo A, then he will overestimate Photo B merely to remain consistent within his newly established frame of reference; this sort of behavior need not be described as "anticipatory," since it is similar to the type of judgment change produced by Tresselt and Volkman (24) in the absence of social interaction. However, as Sherif (19, 20) has pointed out, both anchoring and anticipatory judgments have in common a shift in *S*'s frame of reference. Thus, in some cases, *S* may be anticipating his partner, attempting to "beat him to the punch" (i.e., minimize the anticipated distance between himself and the influence source). In other cases, *S* may still be "anticipating" the experimental situation, which requires a new frame of reference based upon *S*'s previous experience in the situation.

Subsequent Intratrial Influence

The finding that conformity on Photo B is independent of the distance on Photo A raises a question. The present results seem to indicate that distance on the first interaction is a basic determinant only of (*a*) intratrial movement on that first interaction, and (*b*) intertrial movement between the first and second interactions. Can it be concluded, then, that in real life continuous social influence situations, the distance-exposure of an *S* on one interaction has no predictive power in regard to his intratrial behavior in a subsequent interaction on a related issue? This is quite unlikely, when one considers the results of a variety of experiments on "prior reinforcement."

When *S*s, following an initial interaction (or series), have an opportunity to *learn* something about the correctness of their judgments and/or those of their partners (i.e., the influence source), this learning then affects their subsequent judgments differentially (12, 14, 15, 16, 17, 21). For example, if there is some sort of feedback (knowledge of results, or "validation") so that *S* learns he was "right" and/or partner "wrong" on the initial interaction, then subsequent intratrial conformity will be modified. Goldberg and Lubin (9) further found that, when *S* is given some validation information, the distance on the preliminary interactions becomes a significant determinant of subsequent intratrial conformity.

Yet the data from the present study show no such effects: *S*s interacting with a peer partner at different distances on one interaction do not show differential conformity on the next interaction. These findings suggest that, unless *S* can "validate" his own and/or the influence source's judgments in any initial interaction, his intratrial conformity behavior on a second interaction will be independent of the distance he received on the first interaction. Thus, the present study indicates that the variable of distance plays a salient role not only on a particular influence attempt, but in *continuous* social influence processes. (It has also been suggested [5] that distance may be a fruitful concept in the analysis of related influence processes such as psychotherapy.)

SUMMARY AND CONCLUSIONS

Pairs of *S*s were placed in a social influence situation requiring judgments of the number of paratroopers seen in briefly exposed photographs. Two photos were used successively (Cards A and B) after a preliminary warm-up period. On Photo A, distance (the size of the discrepancy between *S*'s judgment and his partner's) was systematically varied by intercepting *S*'s written estimate and replacing it with a bogus estimate at a predetermined distance. After a number of judgments were obtained on Card A, the *S*s were exposed to Card B. On this latter photo, all *S*s were exposed to the same distance. The experiment was replicated with the following major results:

1. At large distances, distributions of the amount of influence on Photo A take on marked positive skewness, the amount of skew increasing with distance.

2. The amount of influence on Photo A is significantly related to distance: (*a*) Movement (the amount of change from initial judgment) increases with distance up to a certain point; there is a slight suggestion (supported by rational consideration) that the curve begins to decrease at extremely large distances. (*b*) Conformity (movement relative to distance) decreases with distance in a monotonic fashion.

3. The amount of influence obtained on Photo A can be increased by sheer repetition of the suggestion, in the absence of any "side arguments," persuasive reasons, or elaborate rationales.

4. When *S*s are permitted a second brief exposure to the stimuli, there is no evidence that *S*s "correct" their previous post-influence judgments toward "reality." In fact, on Card A in particular, the data indicate that *S*s tend to show increased "latent" influence on the re-exposure. However, when *S*s are permitted to count the paratroopers at the end of the experiment, the judgments are no longer related to distance. Part of the earlier influence effects may carry over to this "counting" behavior, depending upon the ambiguity of the stimuli.

5. Considerable amounts of influence are manifested *between* interactions on successive stimuli (intertrial effect). Marked changes are seen in *S*'s initial response to the second stimulus, before any overt attempt to alter his judgment on this particular photo. The intertrial effects are positively related to the distance on the first stimulus. A number of *S*s who were resistant to intratrial influence at large distances on Photo A showed considerable intertrial change.

6. Intratrial influence on the second stimulus was found to be independent of the distance on the first stimulus. It was hypothesized that unless *S* can "validate" his own and/or his partner's judgments in an initial interaction, his intratrial influence behavior on a second interaction will be unrelated to the distance on the first interaction.

REFERENCES

1. ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, (Whole No. 416).
2. FESTINGER, L., GERARD, H. B., HYMOVITCH, B., KELLEY, H. H., & RAVEN, B. The influence

process in the presence of extreme deviates. *Hum. Relat.*, 1952, **5**, 327–346.

3. Fisher, S. The role of expectancy in the performance of posthypnotic behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 503–507.
4. Fisher, S. An investigation of alleged conditioning phenomena under hypnosis. *J. clin. exper. Hypnosis*, 1955, **3**, 71–103.
5. Fisher, S. Plausibility and depth of interpretation. *J. consult. Psychol.*, 1956, **20**, 249–256.
6. Fisher, S., Rubinstein, I., &. Freeman, R. W. Intertrial effects of immediate self-committal in a continuous social influence situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 200–207.
7. French, J. R. P. A formal theory of social power. *Psychol. Rev.*, 1956, **63**, 181–194.
8. Goldberg, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.
9. Goldberg, S. C., & Lubin, A. Influence as a function of prior judgment error. Paper read at American Psychol. Ass., New York City, September, 1957.
10. Hovland, C. I., & Weiss, W. The influence of source credibility on communication effectiveness. *Publ. Opin. Quart.*, 1952, **15**, 635–650.
11. Hovland, C. I., & Pritzker, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.
12. Kelman, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.
13. Kelman, H. C., & Hovland, C. I. "Reinstatement" of the communicator in delayed measurement of opinion change. *J. abnorm. soc. Psychol.*, 1953, **48**, 327–335.
14. Luchins, A. S., & Luchins, E. H. Previous experience with ambiguous and nonambiguous perceptual stimuli under various social influences. *J. soc. Psychol.*, 1955, **42**, 249–270.
15. Mausner, B. The effect of prior reinforcement on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 65–68.
16. Mausner, B. The effect of one partner's success in a relevant task on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 557–560.
17. Mausner, B. Studies in social interaction: I. A conceptual scheme. *J. soc. Psychology*, 1955, **41**, 259–270.
18. Newcomb, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.
19. Osgood, C. E. &. Tannenbaum, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.
20. Schachter, S. Deviation, rejection and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
21. Schein, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 389–395.
22. Sherif, M. *The psychology of social norms*. New York: Harper, 1936.
23. Sherif, M., & Cantril, H. The psychology of attitudes. *Psychol. Rev.*, 1945, **22**, 295–319.
24. Tresselt, Margaret E., & Volkmann, J. The production of uniform opinion by nonsocial stimulation. *J. abnorm. soc. Psychol.*, 1942, **37**, 234–243.

Received April 24, 1957.

THE INFLUENCE OF LANGUAGE-ACQUISITION CONTEXTS ON BILINGUALISM[1]

W. E. LAMBERT, J. HAVELKA,[2] AND C. CROSBY

McGill University

The bilingual person presumably must learn two different verbal responses for each referent, a requirement analogous to learning different responses to the same stimulus in experimental studies of negative transfer. It is of psychological interest, therefore, that bilinguals are so often able to communicate freely in either language without interference from the other. As yet, no adequate psychological theory has been offered to account for bilingualism. In his review of various limited explanations, Weinreich concludes that "a comprehensive psychological theory of bilingualism ought to account for both the effectively separated use of the two languages and for interference of the languages with one another" (9, p. 71). It is the purpose of the present investigation to contribute to such a theory by analyzing conditions that are presumed to affect both the "effectively separated use" of the bilingual's two languages and the "interference of the languages with one another."

As a first step, it may be profitable to apply certain principles derived from paired-associate learning and transfer phenomena to the case of learning two languages.[3] In the experimental study of retroactive inhibition, the learner must first link S_1 to R_1 and, in the later interpolated phase, S_1 to R_2. When alternative responses are associated with a common stimulus, interference between the two associations is the usual consequence. The bilingual is confronted with an analogous learning problem when, for example, the symbols "church" and "*église*" are both associated with the appropriate environmental event.

One general means of minimizing interference in learning situations is to increase the distinctiveness of the two tasks to be learned.[4] In studies of retroactive inhibition, for example, interference is effectively minimized by increasing the degree of learning of original and interpolated tasks or by decreasing the associative similarity between the competing responses. Both these variations enhance the distinctiveness of the tasks. If distinctive secondary cues, such as differently colored backgrounds, are associated with the two tasks to be learned, the functional separation of the tasks should be increased.

Following the logic of interference theory, we can reason that response competition in bilingualism is functionally related to the distinctiveness of contexts in which the bilingual's two languages are acquired; the more "separated" the contexts of acquisition, the less the bilingual interference. Acquisition contexts are considered to be separated when the association between a symbol such as "church" and the appropriate environmental event has consistently taken place in a setting distinct from that in which the association between "*église*" and its corresponding environmental event occurs. On the other hand, in "fused" acquisition contexts, members of a linguistic community can and do use both symbols for any particular environmental event, thereby increasing the interchangeability of the symbols.

Recently, Ervin and Osgood (4) attempted a theoretical interpretation of bilingualism that indicates the potential importance of language-acquisition contexts. They suggest that bilinguals could theoretically acquire "compound" or "coordinate" language systems. The compound system would be developed through ex-

---

[1] This research was supported in part by the Canadian Defence Research Board, Grant Number D77-94-01-10, and in part by N.R.C. Grant A.P. 17.

[2] Now at the University of Western Ontario.

[3] In 1915, Epstein (2) considered bilingualism in terms of such a psychological theory, but did not experimentally test its applicability.

[4] Essentially the same reasoning underlies the "direct method" of language teaching. The direct method theorists maintain that the acquired language should be learned directly from association with environmental events without the mediation of the native language. When learning French as a second language, one should never link "pomme" with "apple" but should link "pomme" directly with the perceived environmental event. In contrast to the "indirect" method which utilizes translation as an aid in acquisition, the direct method is purported to develop "pure" bilingualism, the ability to move from one linguistic and semantic world of experience to another with comparatively little interference.

perience in fused contexts, as with vocabulary training in school, or where the same family members use two languages interchangeably to refer to the same environmental events. The coordinate system would be developed through experience in different linguistic communities where languages are rarely interchanged. They suggest that the alternative symbols in his two languages have a single meaning for the compound bilingual whereas the coordinate bilingual, presumably because of the distinctiveness of his language-acquisition contexts, develops separate meanings for each of his alternative symbols.

Accepting this theoretical position, we hypothesized that experience in "separated" language-acquisition contexts enhances the functional separation of the bilingual's two languages while experience in "fused" contexts reduces the functional separation of the two language systems. We predicted that bilinguals with experience in separated contexts would show comparatively greater semantic differences between symbols in one language and their translated equivalents in the other, more associative independence of translated equivalents in their two languages, and less facility of switching from one language to the other.

METHOD

Classification of Subjects

The *S*s were college or university students with extensive experience in both English and French languages. Since all but four were either preparing for or actually in graduate schools, it can be assumed that the group was relatively homogeneous in intelligence. Each *S* was tested for extent of bilingualism (5, 6) and the 32 who met our criterion for bilingual "balance" will be discussed here. "Balance" implies a high level of second language competence, high enough, in fact, that no dominance of one language over the other is measurable with the technique employed.

Our primary task was to classify *S*s according to their language-acquisition contexts and, to this end, detailed information was obtained from each *S* concerning how, when, and where his languages were acquired. Acquisition contexts were considered *separated* when one language was learned exclusively in the home and the other exclusively outside the home, when one parent consistently used one language and the other parent a different language, or when one language was acquired in a particular national or cultural setting distinct from that in which the second was acquired. Acquisition contexts were considered *fused* when both parents used the two languages indiscriminately, when both languages were used interchangeably inside and outside the home, or when an individual acquired his second language in a school system stressing vocabulary drill and translation and where the first or native language was used as the medium of instruction. This scheme of classification is derived from that used by Ervin and Osgood (4) in their definitions of compound and coordinate systems. The actual classification of *S*s as having separated or fused contexts for acquisition is essentially an arbitrary one in spite of the rather explicit criteria for placement. In five cases, for example, a problem arose in deciding whether to give credit for some experience with separated contexts when an *S* had early experience in a fused context, as when an English *S* acquired French in a school system but later used this language exclusively in France for some period of time. Our decision was to classify such an *S* as separated if he spent at least an entire year using the acquired language either exclusively or primarily. Using this scheme, 23 *S*s were classified as "separated" and 9 as "fused."[5]

It was also possible to further subdivide the "separated" group according to the following line of reasoning. An *S* could be classified as "separated" if he had experience either in distinctive acquisition contexts within a particular national setting, as in Quebec, or in two geographically different cultures, as from living many years in France before coming to America. We predicted that separated contexts based on "bicultural" experiences, as in the latter example, would particularly enhance the distinctiveness of the two contexts. In this case, the environmental events themselves may often not be common for the two sets of symbols. For example, "*église*" could mean a gothic cathedral while "church" could mean a tall wooden building used on Sundays, and "*ami*" and "friend" could take on quite different meanings. A recent study by Ervin (3) demonstrates how such environmental differences affect the perceptual and personality functionings of bilinguals. With these considerations in mind, the "separated" group was subdivided into 15 *S*s with bicultural and 8 *S*s with "unicultural" experiences.

---

[5] In view of our method of selecting *S*s (see 6) where an attempt was made to draw equal numbers of both types, one may safely generalize that it is comparatively easier for bilinguals with separated acquisition contexts to attain bilingual balance.

Behavioral Indices of the Influence of Acquisition Contexts

As a measure of the differences in meanings of concepts in the bilingual's two languages, we made use of Osgood's "semantic differential" (7), a method of scaling stimulus-words on a standard set of meaning dimensions. For example, an *S* is presented with a stimulus-word, such as "house," and is asked to rate it along a seven-point dimension, such as "fast-slow" or "pleasant-unpleasant." In the present case, the stimulus-words were four common English concepts ("house," "drink," "poor," "me") and their French translations. They were presented in a fixed order so that any word was maximally separated from its translated pair. Ten standard dimensions (see 8) were used and their order of presentation was continually changed to eliminate any patterning of ratings. The *S* might rate "house" at Position 3 on the "fast-slow" dimension, and later "maison" might be rated at Position 5 (more slow) on the same dimension, which had been translated into French. Since both stimulus-words and the anchors of each dimension were translated, we can only make statements about differences in meanings of both stimulus-words and dimensional attributes, or their interaction. The average degree of difference between the French and English semantic ratings for the four pairs of stimulus words was determined for each *S* using Osgood's *D* score (7); the larger the *D* score, the greater the semantic differences between translated equivalents in French and English. We predicted that *S*s with experience in separated language-acquisition contexts would show comparatively greater semantic differences between translated equivalents in their two languages.

As a measure of the degree of associative independence of translated equivalents, we made use of the retroactive inhibition design. If an *S* learns list A, then list B, then re-learns list A, one can determine the amount of associative interference imposed by the interpolated material. If the interpolation has no effect on the previously memorized material, then the two can be considered functionally independent. Interference and facilitation from the interpolated material may be thought of as indices of the functional dependence of the first-learned and interpolated tasks.

Each *S* learned a list of 20 commonly occurring English words by the anticipation method. A criterion of 8 correct responses out of 20 was chosen so as to permit improvement following the interpolated task. This was followed by three presentations of a list of 20 three-letter nonsense words. Then the English list was re-presented and the *S*'s retention was measured. Finally, *S*s were asked to relearn the nonsense list, but these data will not be discussed here. This preliminary test gives a measure of the associative interference or facilitation imposed by a meaningless task, one assumed to be semantically independent of the original task. After a rest period, the same procedure was repeated, but this time a second, similar list of 20 English words, learned to the same criterion, was followed by a list of their exact translations in French, presented for three trials. A score was assigned to each *S* which indicated the degree of associative interference or facilitation imposed by the interpolated French material in relation to interpolated meaningless material. The following illustrative case indicates some facilitation from both interpolated tasks but more from the interpolated French. The number of correct responses for the criterion trials are placed in Column A, and the number of correct responses for retention trials are given in Column A′, for the English-Nonsense-English and the English-French-English series:

	A	I	A′	A − A′
Series I:	$E_1 = 10$	*N*(3 trials)	$E_1 = 11$	−1
Series II:	$E_2 = 10$	*F*(3 trials)	$E_2 = 15$	−5

I–II = +4

Positive scores indicate that the interpolated French translations facilitate retention of English words more than do interpolated nonsense words. Negative scores indicate that interpolated French translations interfere with the retention of English words. A zero score would indicate that the interpolated French translations offer no more facilitation than interpolated nonsense words. Following the reasoning outlined above, the closer the scores approach zero, the more associatively independent the translated equivalents. We predicted that *S*s with experience in separated acquisition contexts would show more associative independence between translated equivalents in their two languages than those with experience in fused contexts.

Finally, we predicted that experience in fused as compared to separated contexts would express itself in a greater facility of switching from one language to the other. As a measure of such facility, we made use of a speed of translation test described elsewhere (6). Each *S*, given 20 French and 20 English words to translate, was assigned a score for his average speed of translation from English to French and from French to English. Speed was measured by a voice key in circuit with a chronoscope.

RESULTS AND DISCUSSION

As predicted, those bilinguals who acquired their languages in separated contexts showed a significantly greater difference in meanings of translated equivalents than did those who acquired their two languages in fused contexts (see Table 1). However, when a comparison was made between those who acquired their two languages in geographically distinct cultures (bicultural experience) and those who acquired both languages in separated contexts within one geographical region (unicultural experience), it became clear that bicultural experience accounted for the over-all difference between fused and separated groups. In fact, *S*s with unicultural experience are not essentially different from either of the other two groups with respect to similarity of meanings of translated equivalents. It appears, therefore, that the difference in meanings of translated equivalents is not a simple function of experience in separated acquisition contexts. Experience in bicultural acquisition contexts, where there is a greater likelihood that the actual referents of translated symbols are different, does appear to affect the semantic aspects of the bilingual's two languages.

TABLE 1
SEMANTIC DIFFERENTIAL COMPARISONS

Contexts	N	Mean English-French Profile Separations[a]	Variance		t tests[b]	p values
Fused (F)	9	3.19	0.38	F-S:	2.24	<.05
Separated (S)	23	3.93	1.23	F-B:	2.64	<.03
Bicultural (B)	15	4.22	1.47	F-U:	<1	n.s.
Unicultural (U)	8	3.39	0.33	B-U:	2.13	<.06

[a] Entries are mean *D* scores (7). The larger the mean, the greater the semantic differences between languages.

[b] Where the variances were not homogeneous, p values were determined according to a formula suggested by Edwards (1, p. 273 ff). All t tests in this and the following tables are 2-tailed tests of significance.

TABLE 2
ASSOCIATIVE-INDEPENDENCE COMPARISONS

Contexts	N	Mean Independence Indices[a]	Variance		t tests	p values
Fused (F)	9	+3.44	9.75	F-S:	2.65	<.02
Separated (S)	23	+0.13	11.23	F-B:	2.12	<.05
Bicultural (B)	15	+0.60	10.86	F-U:	2.55	<.02
Unicultural (U)	8	−0.75	12.28	B-U:	<1	n.s.

[a] The closer the scores approach zero, the more between-language associative independence indicated. A constant of 8 was added to each score to eliminate negative values before variances were computed.

Bilinguals with experience in separated acquisition contexts also showed significantly more associative independence of translated equivalents in their two languages than did those with experience in fused contexts (see Table 2). An interpolated list of French translations had about the same influence for the separated group as an interpolated list of nonsense syllables. *S*s in the fused group, on the other hand, clearly benefited from the interpolated French list.

These findings are consistent with those summarized above for semantic differences. Since translated equivalents were semantically more similar for the fused *S*s, the interpolated translations in the learning problem were more like an extension of the original English list. It is logical therefore that these *S*s were helped more by the interpolated French than were the separated *S*s. Since translated equivalents have comparatively different meanings for the separated *S*s, interpolated translations should have shown less associative facilitation and possibly more interference with the original English list, as was the case.

This interpretation has its limitations, however, since the data in Table 2 indicate that there is no difference in degree of associative independence between bilinguals with bicultural and those with unicultural experiences. In fact, both these subgroup means are significantly different from that of the fused group. This finding suggests that experience in separated contexts of either a bicultural or unicultural nature enhances the associative independence of translated equivalents in the bilingual's two languages, as compared to experience in fused acquisition contexts.

TABLE 3
SWITCHING FACILITY COMPARISONS

Contexts	N	Mean Translation Speed[a]	Variance		t test	p values
Fused (F)[b]	7	125.08	112.43	F-S:	<1	n.s.
Separated (S)	23	125.69	288.65	F-B:	<1	n.s.
Bicultural (B)	15	124.59	303.42	F-U:	<1	n.s.
Unicultural (U)	8	127.76	254.62	B-U:	<1	n.s.

[a] Entries, in hundredths of a second, are group means. The larger the mean, the slower the group is in translating into French and into English.

[b] Two *S*s were not available for this test.

Table 3 makes it clear that there were no differences between groups in facility of switching from one language to the other, a finding contrary to predictions. There is no immediately obvious interpretation for this lack of difference between groups, other than the possible lack of sensitivity of our measuring device. Accepting the finding as reliable, one may speculate that *S*s in both groups have had equal experience in translating but possibly different types of experience. The fused *S*s, by definition, have learned their two languages in contexts that encourage translation. On the other hand, a bilingual who has contacts with two separate linguistic communities may often be called on to recount his experiences with members of both communities. In a sense, he becomes the liaison between two linguistic communities and he may learn to use concepts which have meaningful equivalents for members of both communities.

When requested to translate, *S*s in the two groups demonstrate similar facility to switch. In view of the other findings, however, the fused and separated *S*s appear to make different use of the translation process. The fused *S*s made use of the interpolated translations in the learning task and accordingly showed improvement in their retention scores. It is possible, although no attention was paid to this feature in the collection of data, that the separated *S*s were less aware that actual translations were being interpolated than were the fused *S*s. Stated differently, the fused *S*s may have been more set to translate than were the separated *S*s.

In conclusion, we have been able to demonstrate, within the limitations of the small sample of *S*s studied, that language-acquisition context does influence the functional separation of the bilingual's two languages in certain respects. The theory of coordinate and compound language systems has been given empirical support and the defining characteristics of these systems have been extended. The coordinate bilinguals, in contrast to the compound bilinguals, appear to have more functionally independent language systems. If their coordinateness has been developed through experience in culturally distinctive contexts, they will have comparatively different meanings for concepts translated into their two languages. Coordinate and compound bilinguals, however, appear to have equal facility in switching from one language to the other.

SUMMARY

Bilinguals were classified as having learned their two languages in either separated or fused contexts. It was hypothesized that experience in separated as compared with fused language-acquisition contexts comparatively enhances the effectively separated use of the bilingual's two languages.

It was found that experience in separated contexts comparatively increases the associative independence of translated equivalents in the bilingual's two languages. If the bilingual has learned his two languages in culturally distinctive contexts, the semantic differences between translated equivalents is comparatively increased. There was no difference found in facility to switch from one language to the other that can be attributed to contextual influences.

The findings were related to the theory of compound and coordinate bilingual language systems.

REFERENCES

1. EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

2. EPSTEIN, I. *La pensée et la polyglossie.* Paris: Payot, approx. 1915.
3. ERVIN, S. M. The verbal behavior of bilinguals: the effect of the language of report upon the T.A.T. stories of adult French bilinguals. Unpublished dissertation, Univer. of Michigan, 1955.
4. ERVIN, S. M., & OSGOOD, C. E. Second language learning and bilingualism. In C. E. Osgood and F. Sebeok (Eds.), *Psycholinguistics. J. abnorm. soc. Psychol., Suppl.*, 1954. Pp. 139–146.
5. LAMBERT, W. E. Measurement of the linguistic dominance of bilinguals. *J. abnorm. soc. Psychol.*, 1955, **50,** 197–200.
6. LAMBERT, W. E., HAVELKA, J., & GARDNER, R. Linguistic manifestations of bilingualism. (Mimeo.)
7. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49,** 197–237.
8. OSGOOD, C. E., & LURIA, Z. A blind analysis of a case of multiple personality using the semantic differential. *J. abnorm. soc. Psychol.*, 1954, **49,** 579–591.
9. WEINREICH, U. *Languages in contact.* New York: Linguistic Circle of New York, 1953.

Received April 23, 1957.

OBLIGATIONS TO SELF AND SOCIETY IN THE UNITED STATES AND GERMANY[1]

DAVID C. McCLELLAND, J. F. STURR,[2] R. H. KNAPP, AND H. W. WENDT[3]

Harvard University and Wesleyan University

MUCH has been written about the conscious ideals and unconscious motives of both Germans and Americans (3, 4, 5, 7, 8, 9, 11, 13, 14, 16), but little of it has been based on factual evidence systematically gathered for comparative purposes. The present study does not pretend to be complete or definitive because it is based on a very limited sample of upper middle-class male adolescents, but it does attempt (*a*) to see the value systems of the United States and Germany as functioning wholes and to compare them without prejudice as to which is better, and (*b*) to make the comparisons on the basis of actual measurements taken from individual subjects.

While it is fair to state that it was the intention to make the comparison without prejudice, it is also true that no study can be made without preconceptions. So it would be well to state these in advance as explicitly as possible: the general assumption which guided the study was that every society has to strike some kind of a balance between freedom and order, between the wishes of the individual and the obligations of social living, between egocentricism and sociocentricism. Such an assumption has particular value for this study for two reasons. First, two of the best-developed measures of human motivation—n Achievement and n Affiliation—fall easily into one orientation or the other: n Achievement into the egocentric and n Affiliation into the sociocentric orientation. And, second, previous research as well as popular stereotype has tended to picture Americans as "individualistic" (with over-tones of "other-direction" and conformity, to be sure) and the Germans as "sociocentric" (again with a highly individualized sense of honor). So, while this study was designed to be an exploratory inductive enterprise, it was guided by the very general notion that measures should be used that would contrast the two cultures on the obligations to self and to society, although, as the ensuing section will show, many measures were also taken that did not fit precisely into this scheme.

METHOD

Measures

The materials used consisted of a picture-story test of n Affiliation and n Achievement and a questionnaire. The method of deriving n Achievement and n Affiliation scores from brief written stories to pictures has been fully described elsewhere (12, 17). In this study, the pictures used to elicit stories related to achievement were: a male adolescent at a desk, a foreman and worker in a factory, a "father" talking to his "son" with a farm background, and a man working at a desk in his office. The pictures used for measuring n Affiliation were: seven men seated around a table, a mother-son scene (TAT 6BM), a man standing alone looking out through a door, and a figure underneath a lamp post at night (TAT 20). The instructions for eliciting the stories were standard (12) and so far as possible identical in the two countries. The coding of the German stories was done in German using a translation of the English manuals for scoring n Achievement and n Affiliation.

The questionnaire consisted of 38 items of the F-scale type (1) to which the subject (*S*) was to express the degree of his agreement or disagreement on a scale from +3 (for complete agreement) to −3 (for complete disagreement). After the questionnaires were returned, the scale was changed to run from +1 for −3 to 7 for +3, as is common practice, all the results being reported in the latter terms. Three basic orientations influenced the choice of items for inclusion in the questionnaire—the self vs. society dimension already mentioned, previous speculative or research findings on differences between Germans and Americans, and the framework set up by F. Kluckhohn (10) for the comparative study of values. The variety of sentiments covered is illustrated by the fact that at least two items touched on each of the following ideas: achieved vs. ascribed status, loyalty to family and country, emotional control or pride, self-development, kindness and love, community service, affiliative values (friendship), achievement values, other-directedness, man's relationship to nature (e.g., as "over" or "under" nature), the importance of fate and destiny, orderliness and rational planning, and will power. In some cases, a single item was believed to tap more than one of these ideas so that the categories were in no sense conceived as mutually exclusive. In fact, since the categories will play very little part in our analysis, the chief point in mentioning them is to demonstrate how wide a net was cast and to make it easier for subsequent investigators to find obvious areas for comparison that were overlooked.

---

[1] This study was completed when all the authors were connected with Wesleyan University and is based on J. F. Sturr: A comparison of the values of German and American secondary school leaders and non-leaders. Unpublished A.B. thesis, Honors College, Wesleyan University, Middletown, Connecticut, 1955.

[2] Now at the University of Rochester.

[3] Now at the University of Mainz.

There were also three open-ended questions at the end of the schedule. The first asked the student to list his extracurricular activities (clubs, hobbies, sports, etc.), which were then simply counted and coded as involving either group or individual activities. The former included all clubs and team sports (soccer, football, basketball), and the latter individual hobbies (e.g., stamp-collecting) and individual sports (hiking, playing chess or tennis). The second open-ended question asked "What are the three things you would most like to teach your children?" and the answers were coded as being either egocentric (e.g., to do well in school, to appreciate music, to enjoy life) or sociocentric (e.g., to be kind, respect others, to be loyal to one's friends). The final open-ended question asked them to list the "six students in your class who you think are most likely to be leading members of their community 30 years from now." The purpose of this item was to break the students roughly into two groups—one consisting of "leaders" and the other of "nonleaders" at least from the point of view of their peers. In all classes tested, both in Germany and the United States, it turned out that roughly half the *S*s received three or more votes and so this cutting point was used to divide the *S*s into "leaders" and "nonleaders," although it should be clear that the categorization is very crude, and that the term "leadership" is being used in a very restricted sense in this context.

Subjects

The *S*s used were males nearly all between the ages of 16 and 19 (mode 17, median nearly 18 in both countries). Furthermore, all of them belonged to what might be called the educational élite who were preparing to go on for higher education either in a college or university. The nature of the sample was dictated to some extent by the way the German educational system is organized. Beginning at the age of 10, German students who are going on to the university tend to be segregated from students with lesser educational or occupational goals (who go to *Volksschule* or *Mittelschule*), so that a sample of German students tested in the "higher" schools around the age of 18 tend to be a more highly selected group than one would find in the ordinary United States high school senior class. Since it was difficult to get a cross-section of German students at this age, it was decided to limit the comparison to the relatively élite group to be found in German higher schools. Fifty-one students were obtained from a Neusprachliches Gymnasium and 36 from an Altsprachliches Gymnasium, in Kaiserslautern, Germany. To match this sample, *S*s were drawn from college preparatory classes in a high school in Utica, New York ($N = 35$) and from a New England private school ($N = 39$). The private school students were tested to provide a group more nearly comparable to those in the Altsprachliches Gymnasium which emphasized a classical curriculum, and also to make certain that the sample of United States *S*s was as highly selected as the German one. They also make the American sample unrepresentative because private schools exist in large numbers only in the northeastern section of the United States, but the purpose of the study is not primarily to get representativeness of each culture but to contrast the value orientations of comparable young male élites within the two cultures. The testing was done in December, 1954 in the United States and in March, 1955 in Germany.

TABLE 1

Distribution of Occupational Levels of Fathers of Boys in United States and German Samples

Occupational Level	U.S. ($N =$ 70)[a]	Germany ($N =$ 85)[a]
6. Major professional and business executive	32%	24%
5. Minor professional and medium business	26	39
6 and 5 combined	58%	63%
4. Clerical and sales	24	22
3. Skilled mechanical and public service workers	17	14
2. and 1. Semi and unskilled workers	1	1
4 to 1 combined	42%	37%

[a] Information not available on all fathers.

The attempt to obtain boys in the two countries from the same backgrounds was quite successful, as Table 1 shows. As expected, the students came predominantly from the upper middle class and therefore constitute an "élite" group in each country, but the proportion of them from different backgrounds was roughly the same in both countries. The difference between Levels 6 and 5 is not significant and, in any case, is more apparent than real because there were 12 students in the German sample whose fathers were teachers in *Höhere Schule*. According to the coding system, high school teachers in the American sample were classified as minor professionals and placed in Level 5, but the fact of the matter is that such teachers in Germany have a considerably higher status than high school teachers in this country and probably belong at the very least somewhere between occupational Levels 6 and 5. If, to take account of this, half of them were classified in Level 6 or Levels 6 and 5 are combined, as in Table 1, then the distributions of backgrounds of the boys in the two samples become very nearly identical. Furthermore, the percentages of Protestants and Catholics were around 65% and 28% in both groups. In short, any nationality differences found cannot readily be attributed to differences in the socioeconomic level of the two groups of *S*s, or to age differences (since they were matched for age), or to differences in educational plans (since they were all preparing for college), or to differences in religious background.

Results

The mean score for each of the 38 items on the questionnaire was computed for both the

German and American *S*s.[4] About one half of the items revealed differences between the means for the two nationality groups that were significant at or beyond the .05 level. It would be easy and tempting to work out *post hoc* interpretations of these differences, but probably unwise to do so. Even though great care was taken to make the German translations as precise as possible, it is still highly likely that with emotionally toned sentiments such as those used in the questionnaire, the immediate impact of the statement would not be exactly the same on German and American audiences. So at least some of the differences might be attributed to differences in the way the sentiments were expressed. Furthermore, several of the items that on the surface looked as if they should be tapping the same sentiments yielded contradictory results. For example, the Germans agreed much more with Item 29, "There is no satisfaction in any good without a companion" ($p < .01$), but agreed less ($p < .05$) than the Americans with Item 2, "No sane, normal, decent person would ever think of hurting a close friend." Both items deal on the surface with the importance of friendship, but the Germans subscribe more to the former and the Americans to the latter, which incidentally is one of the items drawn (without the final words "or relative") from the F scale on which Germans generally score higher (4). With such variable results, how can one draw general conclusions as to which country values friendship more highly?

For these reasons it was decided to perform a factor analysis on the questionnaire to pick out those items which belonged together in each country. Such a procedure ensures that the items to be combined are chosen beforehand on the empirical basis of covariation, rather than afterwards, in order to explain the particular grouping of results on individual items. Factor analysis also tends to minimize the importance of the translation factor by yielding groups of items that must have been reacted to in somewhat similar ways in the two countries if matching factor patterns can be obtained. A correlation matrix was set up for each country[5] consisting of 36 of the 38 objective items, the two measures of motivation (n Achievement and n Affiliation), plus the number of egocentric virtues mentioned as being desirable to teach children and the total number of activities the *S* listed. Inspection of the matrices showed that practically none of the a priori clusters hung together. That is, items that had been written to tap the same sentiments were not intercorrelated consistently in both countries. So a factor analysis became all the more urgent. It was clear from inspection that certain items showed what appeared to be only random relations with other items. It was therefore decided to include in the factor analysis only those items whose average intercorrelation with the other items departed from zero at the .05 level or better, or those items whose correlations with other items showed a rate of dispersion exceeding the .05 level of probability. However, in order to make matching of factor patterns in the two countries easier, any item was included in both analyses if it met either criterion of significance in *either* country. Rotation to simple structure was performed blindly on the matrix of 19 items chosen by this method and yielded results readily interpretable in terms of two matching factors plus a set of items that shifted from one factor to the other in the two countries, as shown in Table 2. Factor A includes five items in each country that saturate ±30 or more, most of which appear to deal with striving, rationality, and planning. The two most highly saturated items are from the so called v Achievement scale (6), which reflects for the most part the extent to which a person consciously subscribes to achievement sentiments for himself. For simplicity, this factor has been named "rational striving" and perhaps even the one deviant item dealing with the respect due an older man can be thought of as belonging under this heading in the sense that the older man (father figure?) is a kind of symbol or identification figure for people with these sentiments. Factor B contains four items that saturate fairly highly in both countries.

---

[4] These average scores together with the German and American version of each question have been deposited with the American Documentation Institute. Order Document No. 5423 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] We are especially indebted to Ralph Haber who ran off the correlations on a high-speed computer.

TABLE 2

MATCHING FACTOR PATTERNS OF ITEMS FROM A QUESTIONNAIRE ADMINISTERED TO UNITED STATES AND GERMAN MALES OF HIGH SCHOOL AGE

(Saturations less than ±29 omitted)

	Factor A		Factor B	
	U.S.	Germany	U.S.	Germany
Factor A: Rational striving, rationality				
34. I work like a slave at everything I undertake until I am satisfied with the results	43	47	—	—
8. I set difficult goals for myself which I attempt to reach	52	35	—	—
28. Respect is due an older man no matter what kind of a person he is	38	37	—	—
35. A child should never be asked to do anything unless he is told why he is asked to do it	34	38	—	—
6. Nowadays with world conditions the way they are, the wise person lives for today and lets tomorrow take care of itself	−30	−41	—	—
Factor B: "That's life," resignation				
16. There are some people like great artists and musicians who can be forgiven for not being considerate of others, kind to the poor, etc.	—	—	33	56
12. Planning only makes a person unhappy since your plans hardly ever work out anyway	—	—	36	43
23. When a man is born, the success he's going to have is already in the cards, so he might as well accept it and not fight against it	—	—	31	37
13. There is no such thing as a really permanent friendship. Your friends change with circumstances	—	—	40	(25)
C items: (those which shift from Factor A to Factor B) Idealistic concern for others, for "decent" social behavior				
27. A man with money cannot really learn how to behave in polite society if he has not had the proper upbringing	—	48	47	—
2. No sane, normal, decent person would ever think of hurting a close friend	—	38	39	—
7. It is better to go without something than to ask a favor	—	30	31	—
18. There is hardly anything lower than a person who does not feel a great love, gratitude, and respect for his parents	—	29	40	—
Items with unique patterns				
40 Egocentric values mentioned more often than sociocentric as desirable for children	38	−40	—	—
22. It would irritate me very much to have a watch or clock which was off by several minutes every day or so	−32	46	—	—
Items with unique patterns—(*Cont'd.*)				
24. My political opinion is easily swayed by editorials I read	—	−40	—	—
11. When I see a man working in an ordinary unskilled job, I often wonder why he doesn't try to do better in life	—	—	—	−52
15. If you get bad news it is better to hide what you feel and behave as if you didn't care	—	—	—	33
38. A man who gives his life to helping others in trouble is better than a man who spends his life trying to work out schemes so that there will be less trouble in the world	—	—	—	—

It seems to reflect a somewhat passive or resigned attitude toward life with a touch of cynicism, an attitude perhaps best summed up in a shrug of the shoulders and the French phrase "c'est la vie." The interesting thing is that four items switch from being saturated on Factor A in Germany to being saturated on Factor B in the United States. All of them seem to involve a kind of formal concern for the "proprieties," for the abstract code of politeness and obligation to others. As would be expected from this, score on the C items is correlated +.40 ($p < .05$) with Factor A score in Germany but only .06 in the United States. In contrast, the score on the C items is correlated +.28 ($p < .05$) with the score on Factor B in the United States and −.13 in Germany. In other words, explicit concern for "decency" goes with resignation and cynicism in the United States and with rational striving in Germany. The difference is nicely highlighted by the saturations for Item 40, which shows that frequent mention of egocentric virtues is negatively related to Factor A in Germany and positively related in the United States; or to put it more understandably, frequent mention of sociocentric virtues (e. g., loyalty, kindness, honesty) is positively associated with rational striving (Factor A) in Germany but negatively in the United States.

Some of the remaining items with unique factor loadings in the two countries also add to the clear-cut total picture that emerges. For example, perhaps having an inaccurate watch is irritating to Germans high in Factor A because it keeps them from being orderly in fulfilling their obligations, whereas in the United States where orderliness is not a part of the Factor A pattern, being irritable about a watch is not part of the self-picture of the "rational striver." In short, in Germany Factor A is definitely sociocentric: conscientious striving is included in a network of idealistic sentiments relating to decent social behavior. In the United States, rational striving is part of no such framework. On the contrary, sentiments of idealistic concern for "decency" are related to a kind of "un-American" passivity, resignation, and cynicism. In passing, it is perhaps worth noting that two of the C items are from the F scale (one slightly modified) and the other two have an F-scale flavor, a fact which suggests that F-scale-like sentiments may be part of a deviant, rather cynical attitude toward life in the United States, whereas in Germany they are organized into an idealistic pattern dealing with order, efficiency, and striving.

More light on the German-American contrast is shed by Table 3 which shows the mean scores for each nationality group on each of the three sets of items derived from the factor analysis, the two measures of motivation, and the two kinds of extracurricular activities in which the students said they were engaged. The Germans score significantly higher on Factor A ("rational striving") and the C items ("concern for decency") as expected, and as the first line in Table 3 shows, the difference cannot be attributed to a general tendency of the Germans to acquiesce more to items of this sort. That is, since the Germans tend to agree more with all three sets of items included in Table 3, it was necessary to check their acquiescence to all items in the questionnaire to make sure that the differences on these sets of items were not due simply to a stronger tendency for Germans to go along with or acquiesce in emotional sentiments of all sorts (2). No such tendency appeared. The German students further list themselves as being engaged in more outside activities of an individual nature—that is, in activities that do not require group organization or participation.

TABLE 3

MEAN SCORES FOR VALUES, MOTIVES, AND OUTSIDE ACTIVITIES OF U. S. AND GERMAN MALES AGED 16–19 PREPARING FOR COLLEGE

Variable	U.S. (*N* = 74)	German (*N* = 87)	Diff.	*t*	*p*
Acquiescence score: (36 items)[a]					
M	3.74	3.91	−.17	.58	NS
SD	1.12	1.33			
Factor A: (5 items) "Rational striving"					
M	15.45	17.30	−1.85	2.40	<.02
SD	4.83	4.81			
Factor B: (4 items) "That's life"					
M	8.36	9.01	−.65	1.08	NS
SD	3.50	4.09			
C items: (4 items) "Concern for decency"					
M	17.92	20.90	−2.98	4.13	<.001
SD	4.88	4.14			
n Achievement					
M	4.69	2.70	1.99	2.93	<.01
SD	4.45	4.01			
n Affiliation					
M	5.34	5.48	−.14	.24	NS
SD	3.70	3.67			
Group activities					
M	5.18	.97	4.21	12.76	<.001
SD	2.60	1.01			
Individual activities					
M	1.45	2.33	−.88	3.52	<.001
SD	1.78	1.24			

[a] Two items dropped, one because it was too complex to be understood, the other because nearly everyone in both countries agreed fully with it.

The Americans, on the other hand, score higher on n Achievement and much higher in the number of extracurricular *group* activities that they list. Since the difference in n Achievement score is of considerable importance in subsequent interpretation, and since it might have arisen from coding differences or the "pulling power" of the pictures used, it was carefully checked by comparing subcategory frequencies for each of the pictures used. The German students gave less achievement imagery for each of the pictures used, the smallest difference being for the picture of the student and the largest for the "father-son" picture, which may have been somewhat inappropriate for the Germans since the cattle in the background rather suggested to them an American scene. Nevertheless, even among the Americans this picture evoked the least amount of achievement imagery, and even with it omitted, the difference between the two groups is still significant. So it would be hard to attribute the results to differences in the appriateness of picture cues for the two groups.

However, it was found that the German

scorers had used the category, "task imagery,"[6] significantly more often than the American scorers had because, being less familiar with the achievement motive scoring system, they tended, when in any doubt at all, to use this category, whereas the American scorers had been instructed to use it only as a last resort. Consequently, the German mean score might be lower than the American because more real instances of achievement imagery had been classified as doubtful in Germany, which would also mean according to the scoring system that further subcategories could not be scored in such instances. For this reason, a random sample of 46 of the German protocols was rescored in German in the United States by someone with considerable experience in scoring n Achievement in English. The correlation between these scores and the original scores was .70, which is lower than desirable, probably because of the difference in the use of the "task imagery" category. Still, the revised mean score for the German sample was actually *lower* (2.30) than the one given in Table 3, indicating that most of the instances of "task imagery" in the German sample should really have been classified as unrelated to achievement at all. So there seems little doubt that so far as this comparison is concerned, the German students are significantly lower in n Achievement score.[7]

The American boys, then, think more frequently in terms of doing well, a finding that is also supported by the greater frequency with which they mention egocentric rather than sociocentric virtues as desirable to teach to children. "Egocentric" is, of course, not meant in an evaluative sense, but refers to the development of individual capacities (to be intelligent, to appreciate music, to enjoy life, etc.). Unfortunately, the German coders did not force all virtues listed either into an egocentric or sociocentric category as the American coders did, but a comparison of the differences in frequency of mention of the two kinds of virtues in the two countries is still legitimate. That is, in both countries egocentric virtues were mentioned more often, but the difference in favor of them over sociocentric virtues was significantly larger ($t = 2.52$, $p < .02$) in the United States than in Germany.

If the Americans are so concerned with achievement, what keeps them in line? What prevents unbridled individualism? How are they to fulfill their obligations to society? The answer apparently lies in their much greater participation in group activities. They are not brought up on sentiments of obligations to others as the Germans are, but from kindergarten on they regularly participate in many more extracurricular functions of a group nature. In fact, by far the most impressive result in Table 3 is the low number of group activities listed by the Germans (about 1, on the average) as compared with the Americans (about 5, on the average). In these activities the American student must learn a good deal more about getting along with other people and doing things cooperatively, if these clubs are to function at all. If there is any doubt on this point, there is an interesting pair of items correlated positively both in Germany (+.20) and the United States (+.27) on which the two groups of students differ considerably. They are the following:

Mean scores		
U.S.	Germany	
3.30	2.37	#9. The negative opinion of others often keeps me from seeing a movie or play I had planned to attend.
2.84	2.15	#24. My political opinion is easily swayed by editorials I read.

Both groups tend to disagree with these items, but the Germans much more than the Americans ($t = 3.77$, $p < .001$). What is clearly suggested here is the American's greater sensitivity to public opinion, to being guided or directed by others as Riesman has so persuasively put it (15). The American's obligation to develop himself, to achieve, is checked and channelled by the opinion of others in the groups in which he participates so that individualism does not run rampant.

The German is also "individualistic" but in a somewhat different way. He subscribes more than the American to sentiments about rational striving (Factor A) and engages in more individualistic outside activities. Furthermore he agrees more with two other items on the questionnaire which shed light on the nature of this attitude toward the self even though, as evidence, answers to the items do not rank high because they are uncorrelated in either country. The items are as follows:

---

[6] In the scoring system for n Achievement (12) "task imagery" has a value (= 0) between "unrelated" (= −1) and "achievement-related" (= +1) imagery.

[7] Ordinarily such protocols would be scored without knowledge of which group they belonged to but since they were written in different languages, this procedure was not possible in the present case without translating one set into another language. In our view, translation was more objectionable methodologically than the procedure followed here, especially since coding for n Achievement by highly trained scorers is ordinarily done with a re-score reliability of .90–.95, which does not leave much room for unconscious bias to operate.

Mean scores U.S.	Mean scores Germany	
3.28	4.88	#7. It is better to go without something than to ask a favor.
3.09	4.92	#15. If you get bad news, it is better to hide what you feel and behave as if you didn't care.

In both cases the Germans agree much more with the item ($p < .001$) than the Americans do, suggesting that they have more pride in self-control and perhaps think of the self as having the will power to exercise restraint in such cases. How is this kind of self to fulfill its obligations to society? Clearly not by "learning from acting cooperatively in groups" as in America but from a strong sense of social obligation, of one's *duty* to others and of the necessity to put such an obligation above selfish interests.

Evidence was also sought for linkages between the two motives measured and the three groups of value items or the extracurricular activity patterns, but none was found. In correlating the two motive scores with the three types of value scores (Table 3) and with the total number of individual or group activities in each of the two countries (16 different correlations) only one correlation reached the .05 level of significance and it might well have been obtained by chance. The correlation was between n Affiliation and Factor A ("rational striving") in Germany ($r = -.22$). It gains added significance only because the same correlation in the United States was $+.21$ and the difference between the two correlations is highly significant. If anything, this result would appear to mean that spontaneous concern for others (n Affiliation) may make it harder to subscribe to an ideological code of individualistic achievement in Germany, whereas n Affiliation may be positively related to rational striving in the United States because achievement values there are more connected to working with others. Nevertheless, the overwhelming impression of this analysis is that motive scores (as obtained from fantasy) and value scores (as obtained from questionnaires) are not correlated and are essentially independent variables. Consequently, great care should be taken to specify the method of measurement when one speaks about such things as an "interest in achievement," which may refer either to the n Achievement score and its correlates or something like Factor A with its correlates.

Finally, an analysis was made to see if the students nominated by their peers as future outstanding citizens differed in any systematic ways from those not so nominated. For the most part they did not. As far as item or factor differences on the questionnaire are concerned, the two groups of "leaders" did not differ significantly from the "nonleader" groups within either country or from each other between countries. Only one finding of possible future interest emerged: in both countries nominated "leaders" had higher average n Affiliation than "nonleaders" ($p < .05$), a fact which suggests that among peer groups, even across nationality differences, it is the spontaneous interest of individuals in others (n Affiliation) that tends to make them "stand out" as potential leaders in the perception of the group members. Further interpretation of these results is not justified in view of the large number of insignificant findings and the possibility that such a "nomination" procedure may not actually select "leaders" in a very stable or meaningful way.

DISCUSSION

The findings of the present study may be most simply discussed in terms of the following table which has been organized in terms of the double obligation of "the individual in society" to fulfill some of his own needs and also to satisfy some of society's needs:

	Obligation to self	Obligation to society
United States	High n Achievement, concern for egocentric, "self-development" virtues	High participation in group activities, greater sensitivity to the opinions of others
Germany	High v Achievement (concern for rational striving), greater participation in individualistic activities, more emphasis on power of self to exercise restraint	Greater concern over obligations to an idealistic code of decency (governing interpersonal behavior)

Presumably every culture develops more or less patterned solutions to these two problems if it exists for long, and the individual member learns these norms and reflects them in his behavior and attitudes. The point to notice is that according to this analytic scheme, *both* obligations must be at least in part fulfilled by a majority of its members if the society is to function at all. If society should demand too much of its members, it should eventually collapse through *apathy* or *rebellion* because its members can no longer satisfy their needs. On the other hand, if society is neglected through exclusive concern with individual satisfaction, it should eventually collapse through *anarchy*. No one is interested in the welfare of others and society degenerates into the "war of all against all." Between these two theoretical extremes, there should be all sorts of ways of structuring the two sets of obligations which produce working social systems of greater or lesser efficiency. The data of the present study suggest that the scheme is of some value in presenting the somewhat different ways in which morality is structured in the United States and Germany. In the United States a high spontaneous interest in achievement is counterbalanced by much experience in group activities in which the individual learns to channel his achievement needs according to the opinions of others. In Germany a strong sense of self as a striving, controlling entity is offset by an equally strong sense of obligation to a *code* of decency. Interestingly enough, the American "value formula" appears to be largely unconscious or informally understood, as compared to the German one, at any rate. That is, the American boys have higher n Achievement, which is *uncorrelated* with a conscious value placed on striving, and they participate in many group activities without, it is fairly safe to say, having a very explicitly formulated *moral* justification for such behavior. Yet we are arguing that functionally group activities have just such a moral basis in checking excessive individualism. For the Germans, on the other hand, the whole matter appears to be much more consciously worked out. The moral code is more formal and ideological, as other students of German character have observed (14). Individuals are taught and "know" they have a great obligation to the group and strong egos, which are, however, capable of making sacrifices of personal feelings to meet this obligation. In the present study, the German value pattern appears most clearly in the sentiments expressed on the questionnaire, while the United States value pattern appears most clearly in the projective material and the kind of activity the students engage in.

The present study is obviously of very limited scope. It is based on small samples of German and American boys of predominantly upper middle-class status. But the generalizations drawn from it suggest a much wider application to the cultures involved. What evidence is there from previous studies that our conclusions do or do not apply more generally? Has anything been added or subtracted to what is already known about German and American "national character"? In the main, our results are supported by what others have found previously, with one possible change in emphasis to be noted below. Probably the most sensitive clinical or anthropological comparison of the two countries was made by the social psychologist, Kurt Lewin (11), shortly after he immigrated to the United States from Germany. His observations agree at almost every point with our conclusions. He found "probably no other people as interested in *individual accomplishment* . . . as the Americans" (11, p. 277). On the other hand, he calls attention to the flexibility of Americans, to the ease with which they meet strangers, or appear to change their "personality" (opinions and attitudes) to adjust to new situations. In short, he recognized what Riesman has called the *other-directedness* (15) of many Americans which clearly appears in the data of this study.

So far as the Germans are concerned, he described them as carrying more of their "specific individual characteristics to every situation," as having a greater sense of their own identity as separate individuals than Americans do. This kind of "individualism" we have also met in the reports of our German students with the additional note that the self-concept carries with it connotations of personal striving and self-control. Lewin sees the German's sense of obligation to society largely in terms of his tendency to obey orders from higher authority without question. He says, for instance, "In Germany it seems to be the natural right of the adult to rule and the duty of the child to obey" (11, p. 270). In this respect his opinion is supported by a large number of research workers (1, 3, 4, 8, 13, 16). Their point of view is perhaps best summarized by McGranahan as follows: "The German is noted for his unquestioning obedience to authority, his failure to exercise individual responsibility and act on the basis of independent moral judgment" (13, p. 248).

It is at this point that our study differs most from previous ones, and even here the difference is not so much one of fact, but of interpretation. There seems to be no reason to doubt the evidence that Germans are more obedient. The problem arises when attempts are made to explain why they are so obedient, even when such obedience causes great personal suffering or pain to others. Explanations have frequently been given in terms of such motivating factors as "identification with the aggressor," "authoritarian-submission," "sadomasochistic trends," and the like. All such explanations suffer from doing little more than naming the behavior (i.e., Germans obey because they *like to submit*) and from interpreting it in ways which, however acceptable they may be to outsiders (particularly if they dislike Germans), do not make sense to Germans as the rationalizations that at least consciously lead them to obey. The present study does not pretend to get at unconscious factors behind German behavior, but it does shed some light on why they behave as they do *from their point of view.*

To point up the issue as sharply as possible, compare the average agreement with the following item by our United States and German boys:

Mean scores U.S.	Mean scores Germany	
3.50	5.12	#35. A child should never be asked to do anything unless he is told why he is asked to do it.

German boys want explanations significantly more ($p < .01$) than American boys do. On the surface at least, this finding would appear to contradict the notion that Germans are noted for "unquestioning obedience," but one must distinguish between behavior and the attitudes behind it. Suppose we assume that other investigators are correct, that German boys do obey more promptly than American boys, then our problem is to understand why they demand explanations for orders. Here lies the key to the difference between our interpretation and some others. We believe they obey not because they in some sense like obedience and therefore "obey blindly" but because they have some reason, preferably some idealistic reason, for obeying. Without the reason, without the ideal, obedience is meaningless. Hence, we believe, the German emphasis in our data on obligation to an idealistic code of decency.

To make the contrasting interpretations even clearer, consider the following item from McGranahan's study (13) of German and American high school students:

Do you think a boy is justified in running away from home if his father is cruel and brutal?

Three out of four of the American boys said they felt the boy was justified in running away, whereas the German boys split about 50-50. McGranahan is inclined to interpret this finding as supporting the hypothesis that Germans are more submissive and more used to obeying family authority without question no matter how cruel it may be. But our previous analysis suggests that the item does not properly distinguish between what Germans do and why they do it. They may in fact be more submissive to cruel fathers, but not, if our interpretation is correct, because it is good per se to submit but because it is a way of discharging one's duty to an ideology of proper behavior to others (including parents as a very special case). In other words, the German boy may understand the question about running away from a cruel father in terms of whether he is willing to endure personal discomfort to follow the code governing children's obligations to their parents, which is seen by him to be clearly correct in the abstract so that not living up to it for personal reasons seems like a form of selfishness. In fact, the overriding importance of this code of social obligation is frequently stressed in German literature by pointing out that it not only may cause suffering to one's self (thus overriding selfishness) but to others as well. McGranahan and Wayne in an interpretation very similar to ours note that in German plays "we find . . . various murders of beloved persons consciously carried out by sympathetically portrayed characters" (14, p. 454). The characters are sympathetic because they *must* sacrifice even their "nearest and dearest" in the interests of a "higher," more idealistic goal. The same theme appears in a popular German fairly tale, *Der treue Johannes*, in which a father must cut off the heads of his twin sons to fulfill an obligation to a servant whom he has wronged. Apparently, Germans learn in this way that the ultimate test of the sincerity of one's commitment to a code of idealistic obligation is the extreme to which one will go in the sacrifice of personal interests or even the interests of *particular* others. It is in this latter sense that the German concern with decency to others is "ideological" rather than "personal," as McGranahan and Wayne (14) point out.

Outsiders may well wonder how a people so concerned ideologically with what we have called "decency" were also capable of so much cruelty to others under the Nazis. Many have struggled with this problem at deep psychodynamic levels, but a clue to its solution may lie in the two characteristics of the code of obligation just mentioned. First, it seems to be ideological rather than personal. It applies to "others" in the abstract, not to "particular others," so that it doesn't seem so inconsistent to sacrifice particular others in the interest of an abstract code of right conduct designed for the welfare of the whole group. It is perhaps worth noting in passing that the negative correlation in Germany between n Affiliation and Factor A (sentiments of rational striving and decency) may reflect the conflict between the "ideological" and the personal. It may be harder to subscribe to an *ideology* of personal striving and decency if one has much spontaneous interest in *particular* others (n Affiliation) in a country where particularistic interests are often thought of as conflicting with one's larger obligations.

Second, cruelty may actually appear as the validating sign of the strength of one's commitment to the code, as in *Der treue Johannes*. One must show that he can "take" even doing horrible things to *prove*[8] his commitment to a higher end. Since on this point McGranahan and Wayne's study came to the same general conclusion based on pre-Nazi data as ours does on post-Nazi data, it appears that what the Nazis did was to play up a tendency to idealistic sacrifice of personal interests, even of affiliative ones, always present in German character.

In the main then, our conclusions are supported by other students of German and American character. We differ chiefly in seeing German obedience as a response that at least from the internal point of view is dictated not by a liking for obedience but by a strong belief in the importance of sticking to an abstract, ideological code of right conduct in society. As confirmation of this alternative interpretation, McGranahan and Wayne (14) found that obedience is *not* always good in German plays: but the one time rebellion is legitimate is when it is in the interest of some higher good of society as a whole. In short, it is not obedience per se which is good. It is seen as instrumental to an ideology of right social behavior that may on occasion demand rebellion.

SUMMARY

A questionnaire covering various value sentiments and a picture-story test for measuring n Achievement and n Affiliation were administered to 87 German and 74 United States male students preparing for college who had been carefully matched for age (around 18 years) and socioeconomic background (largely upper middle class). A factor analysis of the questionnaire results yielded two factors that could be matched in the two countries and a third set of items that shifted from one factor to the other across countries. The "switch-over" items which dealt largely with decency and obligation to others were associated with Factor A ("rational striving") in Germany and were more strongly accepted by the Germans than by the Americans, as were also the Factor A items. In the United States, the switch-over items were associated with Factor B ("resigned cynicism") suggesting that acceptance of statements about strong obli-

---

[8] One may even wonder, somewhat paradoxically, if the Germans are forced, functionally speaking, to emphasize self-sacrifice so much just because they have such a strong sense of self. Perhaps the problem is less acute for the American whose sense of personal identity is less well-developed in the sense that he talks more freely to others (Items #7, #15) and is more easily guided by others (Items #9, #24).

gations to others tend to be associated in the United States with a kind of deviant "un-American" passive resignation. Further analysis showed that the American students were significantly higher than the Germans on n Achievement, on participation in group activities, and in the extent to which they said they were guided by the opinions of others. The Germans for their part reported a greater number of individualistic outside activities and subscribed more to sentiments picturing the ego as independent of others. These results were interpreted to mean that in the United States obligation to self appears as a strong unconscious need for Achievement, and obligation to society as participation in many group activities, and that in Germany obligation to the self appears as a stress on the ego as a separate, unique, willing entity and obligation to society as a moral imperative to place obligation to the "impersonal other" as an abstract code above particularistic, "selfish" considerations. An attempt is made to interpret German "authoritarianism" in terms of duty to such an abstract code rather than in terms of the more usual, somewhat unanalytical and ethnocentric assumption that Germans "like" to dominate and submit more than other people.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritian personality.* New York: Harper, 1949.
2. BASS, B. M. Authoritarianism of acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
3. BRICKNER, R. *Is Germany incurable?* Philadelphia: J. B. Lippincott, 1943.
4. COHN, T. S., & KARSCH, H. Administration of the F-scale to a sample of Germans. *J. abnorm. soc. Psychol.*, 1954, **49**, 471.
5. COMMAGER, H. S. *America in perspective.* New York: Random House, 1947.
6. DECHARMS, R., MORRISON, H. W., REITMAN, W. R., & MCCLELLAND, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955.
7. DE TOCQUEVILLE, A. *Democracy in America.* (1833). New York: Knopf, 1945.
8. FROMM, E. Hitler and the Nazi authoritarian character structure. In T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1947.
9. GORER, G. *The American people.* New York: Norton, 1948.
10. KLUCKHOHN, F. R. Dominant and substitute profiles of cultural orientations: Their significance for the analysis of social stratification. *Social Forces*, 1950, **28**, 376–393.
11. LEWIN, K. Some socio-psychological differences between the United States and Germany. *Charact. and Pers.*, 1936, **4**, 265–293.
12. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
13. MCGRANAHAN, D. V. A comparison of social attitudes among American and German youth. *J. abnorm. soc. Psychol.*, 1946, **41**, 245–257.
14. MCGRANAHAN, D. V., & WAYNE, I. German and American traits reflected in popular drama. *Hum. Relat.*, 1948, **1**, 429–455.
15. RIESMAN, D., WITH GLAZER, N., & DENNEY, R. *The lonely crowd.* New York: Doubleday, 1953.
16. SCHAFFNER, B. *Fatherland: A study of authoritarianism in the German family.* New York: Columbia Univer. Press, 1948.
17. SHIPLEY T. E., JR., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.

Received April 29, 1957.

SEX DIFFERENCES IN PROBLEM-SOLVING PERFORMANCE AS A FUNCTION OF ATTITUDE DIFFERENCES[1]

GLORIA L. CAREY[2]

University of Nebraska

PROBLEM solving is one of several areas of intellectual functioning in which sex differences in performance have been noted (1, 3, 5, 6). Usually these differences have been reported as incidental to the major findings, and very few studies have been designed specifically to test sex differences in reasoning or problem-solving ability. Sweeney (7), however, did perform a series of experiments in which he controlled variables he thought might be responsible for the previously noted masculine superiority in problem solving. Equating men and women on aptitude (mathematical and verbal), visualization, mechanical, comprehension or mathematical background did eliminate sex differences in performance on certain types of problems, but a striking number of highly significant sex differences remained. Sweeney concluded that these remaining differences occurred on problems that required restructuring or involved a general reasoning factor.

The present investigation began with the idea that some of the sex differences in problem solving could be accounted for by differences in nonintellectual factors. The question that needed to be asked is: How well do women like to solve problems? If problem solving is an activity preferred by men and disliked by women, there exists a good possibility that sex differences in performance are actually a reflection of sex differences in attitude.[3]

Such observations as the relative lack of women in the scientific professions, feminine expressions of inadequacy and dislike in problem-solving experiments, and the cultural stereotype of *man* as the "reasoning animal" point toward a negative attitude toward problem solving and the reasoning process, and make the following hypothesis plausible: sex differences in problem-solving performance which are not the result of differences in general intelligence, special aptitudes, or information, are attributable to differences in attitude toward problem solving.

In order to test this hypothesis, studies were designed to answer the following specific questions:

1. Can a scale be constructed that measures attitude toward problem solving?
2. Is there a sex difference on such a scale?
3. Is problem-solving attitude related to problem-solving performance?
4. Is an attempt to change problem-solving attitude experimentally followed by a change in performance?
5. Do women respond more favorably than men to an attempt to improve their attitudes? If women have an initially more unfavorable attitude than do men, there should be more opportunity for improving the attitude of women than of men, and hence one might expect a greater relative improvement in their performance.

Affirmative answers to the first three questions would give support to the major hypothesis, but it was felt that such answers to the fourth and fifth questions would provide even more convincing evidence. In addition, a scale throws little light on the question of whether performance depends on attitude, or whether the reverse is also true—that attitude depends on performance. Therefore, the main experiment was designed to answer Questions 4 and 5, which concern a deliberate manipulation of the independent variable (attitude) and measurement of its effect on the dependent variable (performance).

---

[1] Condensed from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Stanford University, 1955. The writer is indebted to Drs. D. W. Taylor, P. R. Farnsworth, R. R. Sears, and Olga McNemar for their help and advice. She also wishes to thank Drs. Yvonne W. Brackbill, D. L. LaBerge, and R. J. Rhine for their assistance in conducting experimental sessions. The investigation was carried out under Project NR 150-149 supported by Contract N6onr 25125 between Stanford University and the Office of Naval Research. Permission is granted for reproduction, translation, publication, use, and disposal of this article in whole or in part by the United States Government.

[2] Now at Swarthmore College.

[3] By "attitude" the writer is referring simply to liking or not liking to solve problems, or to the valence which problem solving has for the individual. This does not mean that it is possible to distinguish clearly between attitude and other nonintellectual factors in problem-solving success, such as motivation and interest.

METHOD

Attitude scale. The Likert scaling method was applied to 63 statements which, in the opinion of several investigators working on a problem-solving research project, might be related to attitude toward problem solving. (Example: "I would rather have someone tell me the solution to a difficult problem than to have to work it out for myself.") There was no attempt deliberately to construct items which would show a sex difference in response. On the basis of an item analysis of the responses of 32 men and 100 women in an elementary psychology course at Stanford University, 36 items were selected that differentiated between high and low scorers at the .05 level or better. The 36 items were then split on the basis of similarity of content and form into two groups of 18 items each, which were referred to as Forms A and B of the attitude scale. For this first sample, the intercorrelation between the two forms administered at the same time was .94. The intercorrelation for a second sample of 59 women and 50 men was .83.

Attitude modification. In order to study the hypothesized effect of attitude on performance, the following experiment was designed so as to measure attitude and performance both before and after an attempt was made to modify attitude. Sixteen groups of six subjects (*S*s) each met together for a total of three hours, one hour on each of three successive days. Each group was composed of three men and three women—making a total of 48 men and 48 women. All *S*s were members of an elementary psychology class at Stanford University, and the majority of them were sophomores. During the first hour *S*s were given one form of the attitude scale and also a series of ten problems to solve, with 40 minutes working time allowed for completion of the problems. The problems were ones for which a significant sex difference had appeared in an earlier experiment (7) even though the sexes had been matched on variables thought to be related to performance differences. The second hour consisted of an informal group discussion of some of the factors involved in problem-solving success, with the implicit aim of (*a*) conveying an appreciation of the relationship between attitude toward problem solving and problem-solving success and (*b*) encouraging each *S* to explore and recognize his or her own attitude toward problem solving. During the third hour, the alternate form of the attitude scale and a second, comparable set of ten problems were administered. Since the two sets of problems were not exactly equal in difficulty, the order of presentation was reversed for half the *S*s. At no time was a reference made to the real purpose of the experiment, nor was the subject of sex differences ever introduced. As far as the *S*s were concerned, the first and third hours of the experiment were used to standardize the problems and scales, and the second (discussion) hour was used to get their suggestions for future problem-solving research.

There were four experimenters (*E*s), two male and two female, each of whom worked with four groups of six *S*s each. They administered the attitude scales and problems, and led the group discussions. They also observed two demonstration discussions by the writer before leading discussions of their own. The use of four *E*s allowed more generality to the conclusions than would be possible with one *E*. Two *E*s of each sex also provided an opportunity to test whether a significant interaction occurs between sex of *E* and sex of *S*.

If a relatively greater improvement in performance were obtained for women than for men, the possibility would exist that women benefited more from the practice involved in working the first set of problems than did men. As a group, women may have had less experience with the types of problems used and hence may show a greater practice effect. For this reason, control groups of 28 women and 20 men were used, the procedure being the same as that for the experimental *S*s, with the exception that the discussion period was eliminated for the control groups. In each control group, the order of administration of the problems and attitude scales was reversed for half the sample, as it was for the experimental *S*s.

RESULTS

The results obtained from the attitude scale and from the attitude modification experiment are summarized with respect to each of the five questions raised earlier.

1. *Can a scale be constructed that measures attitude toward problem solving?* A reliable measuring instrument was constructed. Some internal evidence of validity was provided by the fact that the items that referred directly to interest in or frequency of problem-solving activity were among the most highly differentiating items in the item analysis. Additional evidence of validity was provided by the fact that attitude scores and performance scores were positively related (see Question 3), which indicates that the people who say they like to solve problems or are interested in related activities also solve the most problems. It would be desirable to know the frequency with which *S*s with high attitude scores actually seek or indulge in problem-solving behavior, but even without such a behavioral criterion of validity the contents of the scale and measures of validity cited above provide some evidence that the scale is actually measuring attitude toward problem solving.

2. *Is there a sex difference on such a scale?* Men received significantly higher scores on the attitude scale than did women. In Table 1 is a summary of the data for the two samples used.

3. *Is problem-solving attitude related to problem-solving performance?* Correlations were computed between the attitude scale scores and problem-solving scores for men and women separately in both the prediscussion and the postdiscussion testing sessions. The eight correlations, seven of which were in the predicted direction, are given in Table 2. The fact that

TABLE 1

SEX DIFFERENCES IN ATTITUDE

Sample	N	Mean	SD	t	p
First Sample					
Form A					
Men	32	63.12	8.83	3.08	.01*
Women	100	57.61	8.71		
Form B					
Men	32	58.19	9.68	2.74	.01*
Women	100	53.17	8.70		
Second Sample					
Form A					
Men	50	55.88	7.56	1.86	.04*
Women	59	52.68	9.86		
Form B					
Men	50	60.08	7.10	2.39	.01*
Women	59	56.42	8.48		

* An asterisk in this and the following tables indicates that a one-tailed test of significance was made, since the direction of the difference was predicted.

TABLE 2

CORRELATIONS BETWEEN ATTITUDE AND PERFORMANCE SCORES: EXPERIMENTAL GROUPS

Session and Sex	N	Attitude A and Problems I		Attitude B and Problems II	
		r	p	r	p
Prediscussion					
Men	24	.57	.001*	.24	.13*
Women	24	.29	.08*	−.08	—
Postdiscussion					
Men	24	.60	.001*	.32	.07*
Women	24	.20	.18*	.29	.08*

only two were highly significant may have been due in part to the small N for each subgroup.

4. *Is an attempt to change problem-solving attitude experimentally followed by a change in performance?* Performance data for the pre- and postdiscussion testing sessions are given in Table 3. A significant improvement in the performance of women indicates that the question may be answered in the affirmative for the female *S*s.

5. *Do women respond more favorably than men to an attempt to improve their attitudes?* In order to answer this question, a two-way analysis of variance design was used to analyze the difference scores, with classification according to *E* and sex of *S*.[4] Table 4 is a summary of the analysis of variance results. Difference scores were calculated by subtracting each *S*'s problem-solving score for the prediscussion session from that for the postdiscussion session. Since the relative improvement of female *S*s (mean difference score of women minus mean difference score of men) was 1.25 and 1.22 for one female *E* and one male *E*, respectively; and −.08 and .22 for the other female *E* and male *E*, respectively, there was obviously no significant interaction between sex of *E* and sex of *S*. Hence, the classification according to sex of *E* was not used.

TABLE 3

DIFFERENCES IN PROBLEM-SOLVING SCORES BETWEEN FIRST AND SECOND SESSIONS: EXPERIMENTAL GROUPS

Sex	N	First Session		Second Session		Diff.	t	p
		Mean	SD	Mean	SD			
Men	48	5.13	2.04	5.33	1.99	.20	.48	—
Women	48	3.78	1.81	4.63	1.73	.85	2.30	.02*

TABLE 4

ANALYSIS OF VARIANCE OF PROBLEM-SOLVING DIFFERENCE SCORES: EXPERIMENTAL GROUPS

Source	Sum of Squares	df	Var. Est.	F	p
Sex difference	10.24	1	10.14	2.86	.05*
Exper. difference	40.45	3	13.48	3.80	.03
Interaction	8.44	3	2.81	.79	—
Individual differences	312.51	88	3.55		

The sex difference in amount of improvement in performance was significant, women responding more favorably than men to the attempt to improve problem-solving attitude. An additional finding was that the sex difference in amount of improvement varied according to *E*, a fact that indicates the need for using several *E*s to yield results of some generality.

The change in performance of the control groups was not significant for either sex ($t =$ .77 and .28 for men and women, respectively). This finding indicates that the amount of practice represented by the first session did not result in a significant improvement in performance.

As shown in Table 5, the improvement of the

---

[4] As a check on the consistency of the improvement or lack of improvement for the women in the groups led by the four *E*s, Bartlett's test for homogeneity of variance was applied to the variance estimates of the four *E*s' groups. Since chi square = 4.49 ($p = .19$), the assumption of homogeneity of variance was not rejected. That is to say, the groups did not differ in variability.

TABLE 5

DIFFERENCES BETWEEN MEAN DIFFERENCE SCORES FOR EXPERIMENTAL AND CONTROL GROUPS

Sex	Mean Difference		Diff.	t	p
	Experimental	Control			
Men	.20	.55	−.35	.78	—
Women	.85	−.14	.99	2.11	.02*
Difference between men and women			1.34	2.06	.02*

experimental as compared with the control groups was significantly larger for women than for men. This finding further confirms the conclusion reached earlier on the basis of the analysis in Table 4: women showed significantly more improvement as a result of the discussion than did men.

A comparison was also made of the pre- and postdiscussion attitude scores. Contrary to expectation, women did not show a relatively greater change in attitude scores than did men, nor did the scores improve significantly for either sex following the discussion. Male and female *S*s in the control groups also showed no significant improvement in attitude scores and a comparison of the change for experimental versus control *S*s revealed no significant differences.

DISCUSSION

The answers to the five questions support the hypothesis that previously reported sex differences in problem-solving preformance are attributable in part to sex differences in attitude toward problem solving. An explanation of this difference in attitude probably lies in our culture, which still adheres to some extent to the view that "Woman is the unreasoning animal and pokes the fire from on top."[5] It is hypothesized that men tend to be more favorably disposed toward problem solving because they have been encouraged in this attitude and because acknowledgment of an unfavorable attitude is a denial of one aspect of the masculine role. Some support for this view is shown in the relationships found between attitude toward problem solving and mathematical and verbal aptitude (2). The correlations for women's aptitude and attitude scores were highly significant for both aptitude measures, but there was no relationship between men's aptitude and attitude scores—a finding which indicates that men tend not to admit to a dislike for problem solving even when they are lacking in intellectual aptitude.

Although a sex difference in improvement of problem-solving performance was the major prediction, it was also predicted that there would be a sex difference in improvement of attitude scores. The lack of any improvement in attitude scores warrants further consideration.

One explanation for this finding is that the scale was not sensitive enough to reflect the changes in attitude that occurred, since a number of the items referred to interests and activities with respect to which *S*s would have had no opportunity to alter their behavior before the second testing session (e.g., "I enjoy being the score keeper when playing bridge or canasta"). Another explanation for the lack of improvement in attitude scores is based on the occurrence of catharsis during the discussion. In this permissive situation *S*s were encouraged to explore their attitudes toward problem solving—whether positive or negative—with the result that they felt quite free to express negative attitudes. The cathartic effect of discussion may have been such as to result in a freer, more realistic expression of attitude.

A final possibility, of course, is that the discussion did not modify attitude and that other factors accounted for the improvement in women's performance. Practice in solving the first set of problems might result in such an improvement if women were more affected by the practice than were men, but this factor was controlled by the use of female *S*s who had a similar exposure to the problems without the intervening discussion. The fact that the control *S*s did not improve in performance suggests that the effect of solving the first set of problems was to counterbalance the positive effects of practice by reinforcement of the existing negative attitudes. An examination of the discussion procedure indicates that motivations such as sex rivalry, increased level of aspiration, or a desire to please the discussion leader do not appear to account for the improvement in women's performance. At no time did *S*s in the discussion groups comment on sex differences in problem-solving performance, and the *E*s were under instructions not to refer to such differences or to the real purpose of the discus-

[5] Quoted by William James, in a discussion of why women are relatively poor at reasoning (4, p. 368).

sion. In fact, *S*s did not know that they would be given more problems to solve after the discussion session, although the remark was occasionally made at the end of the discussion that if *S*s had a chance to work on more problems, they thought they would do a better job of solving them. No scores or answers were given out during the experiment so that change in level of aspiration is also not a very likely possibility. By the use of both male and female *E*s, interaction between sex of *E* and sex of *S* could be determined, but the results indicated that such interaction did not occur. Since *E*s made no attempt to exhort *S*s to improve their attitudes or performance, it is unlikely that female *S*s improved in performance because they were motivated to please *E*, although this possibility cannot be ruled out entirely.

Consideration of possible alternative explanations for the significantly greater improvement in women's performance leaves change in attitude as a tenable one even though the attitude scale itself was probably not sensitive enough to reflect a change.

A final comment may be made on the fact that there were marked differences among *E*s in the relative improvement of female *S*s. The *E* difference was not due to an interaction between sex of *E* and sex of *S*, and the discussion leaders were of approximately the same age and level of academic attainment. Since the discussion procedure was also fairly well standardized from one group leader to another, it would appear that the differences were related to personality characteristics of the *E*s.

SUMMARY

The study was designed to test the hypothesis that sex differences in problem-solving performance which are not the result of differences in general intelligence, special aptitudes, or information, are attributable to differences in attitude toward problem solving. First, a scale designed to measure attitude toward problem solving was constructed. Then the attitude scale and a series of ten problems were administered to experimental groups consisting of three men and three women each. The group members then participated in a discussion, with the disguised intent of promoting a more favorable attitude toward problem solving. In a postdiscussion session, *S*s were administered an alternate form of the attitude scale and a second set of problems, comparable to those used in the prediscussion session. A control group that did not participate in the discussion was also used.

The data were analyzed according to their bearing on five questions related to the major hypothesis. 1. Can a scale be constructed that measures attitude toward problem solving? 2. Is there a sex difference on such a scale? 3. Is problem-solving attitude related to problem-solving performance? 4. Is an attempt to change problem-solving attitude experimentally followed by a change in performance? 5. Do women respond more favorably than men to an attempt to improve their attitudes?

The attitude scale gave some internal evidence of validity, and men were found to have significantly higher scores on the scale than women. Also, attitude scores were found to have some positive relationship to performance scores. After the group discussion, designed to induce a more favorable attitude toward problem solving, women showed a significant improvement in performance, while men showed no change. An examination of some possible ways of accounting for the relatively greater improvement of women's postdiscussion performance indicated that improvement of attitude is at least a tenable one. In general, the findings support the hypothesis of sex differences in problem solving as a function of sex differences in attitude. An explanation of such attitudinal differences in terms of cultural differences for men and women is offered.

REFERENCES

1. BILLINGS, M. L. Problem solving in different fields of endeavor. *Amer. J. Psychol.*, 1934, **46**, 259–272.
2. CAREY, GLORIA L. *Reduction of sex differences in problem solving by improvement of attitude through group discussion.* Stanford: Dept. Psychol., Stanford Univer., Tech. Rep. No. 9, 1955 (Contract N6onr 25125).
3. GUETZKOW, H. An analysis of the operation of set in problem-solving behavior. *J. gen. Psychol.*, 1951, **45**, 219–244.
4. JAMES, W. *Principles of psychology.* Vol. 2. New York: Henry Holt, 1890.
5. MAIER, N. R. F. Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, **35**, 349–360.
6. SAUGSTAD, P. Incidental memory and problem solving. *Univer. Chicago Psychometr. Lab. Rep.*, 1951, No. 66.
7. SWEENEY, E. J. *Sex differences in problem solving.* Stanford: Dept. Psychol., Stanford Univer., Tech. Rep. No. 1, 1953 (Contract N6onr 25125).

Received April 29, 1957.

PERSONALITY AND IQ CHANGE[1]

JEROME KAGAN, LESTER W. SONTAG, CHARLES T. BAKER,
AND VIRGINIA L. NELSON

The Fels Research Institute, Antioch College

RESEARCH on mental development during the last twenty years has indicated that a child's IQ score does not necessarily remain constant with age (2, 3, 4, 10). Several reports (9, 10, 12) suggest that changes in environmental conditions can depress or raise IQ level and it is sometimes implied that these changes may be explained by recourse to personality variables. The purpose of this paper is to demonstrate that changes in IQ during childhood are correlated with certain personality predispositions as inferred from projective test data. The personality variables under study include (*a*) need for achievement, (*b*) competitive strivings, (*c*) curiosity about nature, and (*d*) passivity.

Performance on an IQ test is assumed to be a function of at least two major variables; the variety of skills and abilities the person brings to the test situation and his motivation to perform well on the test (2, 6). Since the IQ scores of some children change markedly during the school years, it seems plausible to assume that those children who show marked increases in IQ have a very strong motivation to acquire or develop the various intellectual skills tapped by an IQ test and to perform well in a testing situation. It is suggested that need for achievement, competitive strivings, and curiosity about nature motivate the acquisition and improvement of cognitive abilities and by so doing facilitate increases in tested IQ.

The social environment often awards praise and recognition for intellectual accomplishment, and school age children with a high need for achievement might seek to gratify this need through intellectual activity. Thus it was predicted that children showing marked increases in IQ would produce more achievement imagery on the TAT than those with minimal gains in IQ.

Secondly, the school environment emphasizes competitive intellectual activity, and children with strong competitive needs would be highly motivated to acquire the intellectual skills which result in successful competition with one's classmates. Thus it was predicted that children showing IQ gains would show more competitive strivings than children displaying minimal gains in IQ. In choosing an index of competitive strivings, besides the related measure of TAT achievement fantasy, it was decided to use aggressive content on the Rorschach. The bases for this choice rested on the assumptions that (*a*) incidence of aggressive imagery reflected degree of aggressive motivation and (*b*) competition was a socially accepted form of aggressive behavior. For in competition, as in aggression, the child desires to defeat another individual and assert his superiority over him. The population of children in this study is predominantly middle class and apt to place strong inhibitions on direct, overt expression of aggression. Therefore, there would be a tendency for the individual with high aggressive motivation to seek socially accepted channels for aggressive expression such as competitive activity with peers. Thus it was predicted that children showing IQ gain would report more Rorschach aggressive content than those with minimal gain because of their greater competitive predisposition.

A third motive that might facilitate a child's acquisition of knowledge and skills in dealing with the environment could be curiosity about nature. Interest in birth, death, sexual anatomy, and other processes of nature is a frequent phenomenon in young children. It is suggested that the more intense this curiosity the greater the motivation to acquire the habits which would gratify this motive. Since reading, questioning, and manipulating the environment are effective behavioral methods of gratifying one's curiosity, it might be expected that the highly curious child would be more likely to develop these skills and therefore apt to gain in IQ score. The TAT measure used to evaluate curiosity was presence of themes of interest in nature and its phenomena. For the Rorschach, it was hypothesized that concern with the body might reflect, in part, heightened interest in natural processes, and it was suggested that anatomy content might be more frequent for children who showed marked IQ gains than for those with minimal increases in IQ. It is recognized that many clinical psychologists regard anatomy content in adults as indicative of psychopathology. This study is concerned with the correlates of IQ gain rather than psychopathology, and it is not implied that children who show increases in IQ are completely free of conflict. Secondly, it was felt that the determinants of anatomy content for children might be different from those which produce this content in adults.

---

[1] This investigation was supported in part by a research grant (PHS M 1260) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The writers wish to thank Dr. Seymour B. Sarason for his critical reading of the manuscript.

A final prediction dealt with the predisposition to behavioral passivity. The children who show IQ gains have been characterized as having high need achievement, competitive strivings, and curiosity about the environment. This constellation of motives implies that when these children are confronted with a problem, they would have a tendency to attack and attempt to solve the problem rather than withdraw from the situation or seek help. On this basis, it was predicted that children who showed IQ gains would be less likely than those with minimal IQ increases to characterize their TAT heroes as passive in attitude or behavior.

The Fels Research Institute is uniquely equipped to test these ideas about IQ change since it has continuous longitudinal information on the development of a sample of normal children. These data include intelligence and projective tests, observations of the children, and reports on the parent-child interaction. In a recent study, Sontag, Baker, and Nelson (11) related personality information on a sample of children with changes in IQ and found that those children who showed marked increases in IQ were rated as more competitive, more likely to display self-initiated behavior and less passive than those who showed decreases in IQ. The TAT and Rorschach protocols were not utilized in making these personality ratings, and the results from this study served as a major stimulus for the present investigation.

METHOD

A sample of 140 Fels subjects (*S*s), 70 of each sex, were chosen for study because a fairly complete record of test information was available on them. From ages 2½ to 6, the Stanford-Binet intelligence test (1916 or 1937 revision) was administered to most *S*s twice yearly, on their birthdays and six months after their birthdays. From ages 6 to 12, most *S*s received alternately Form L or Form M of the 1937 revision annually on or near each *S*'s birthday. All of the tests were administered by one of the authors (VLN). The mean IQ of the Fels population is near 120, with standard deviation varying from 14 to 20 IQ points.

In order to obtain groups of *S*s who showed the most change in IQ score from ages 6 to 10, a smoothed longitudinal plot of each *S*'s IQ was prepared by averaging the mean of three consecutive test scores around each age. This procedure is explained in detail in other reports (1, 10, 11). This technique tends to eliminate erratic variations in IQ and hopefully furnishes a more valid measure of IQ changes. Then each *S*'s smoothed IQ at age 6 was subtracted from his smoothed IQ at age 10, and this distribution of differences, positive if *S* gained in IQ and negative if *S* lost in IQ, was divided into quartiles. This report deals with the projective test information on those *S*s in the two extreme groups; those who increased and those who decreased the most in IQ score. These will be called Group A, the IQ ascenders, and Group B, the IQ descenders, respectively. There was no significant difference between the mean IQ of the two extreme quartiles at age six, the means being 119 and 116 for Groups A and D respectively. The average amount of increase in IQ for Group A was larger (plus 17 points) than the corresponding decrease for the members of Group D (minus 5 points) and while 46 per cent of Group D lost five or more points, every child in Group A gained 10 or more points during the years 6 through 10. The mean IQ of the entire sample of 140 tends to increase slightly from ages 6 to 10, probably as a result of practice effects with the same test. Since every *S* in Group D showed a decrease in IQ, it might be inferred that the members of Group D did not benefit from practice and familiarity with the test, and it is probably more accurate to view Group D *S*s in this light rather than as *S*s who showed marked decreases in IQ score.

The projective tests used in the analysis were the Rorschach and selected TAT pictures. Two factors governed the choice of the TAT cards which were analyzed. Because the protocols were gathered over a period of years, there was not complete comparability for all *S*s for the number of cards administered. Secondly, the specific hypotheses of the study dictated the cards chosen for analysis and Cards 1, 3 BM, 3 GF, 5, 6 BM, 12 F, 14, and 17 BM were selected for analysis. The age at which the TAT protocols were administered ranged from 8–9 to 14–6 with median at 11–6 and 80 per cent of the protocols obtained between the ages of 11 and 12. The age at which the Rorschachs were administered ranged from 6–5 to 13–6 with median at 10–5 and 63 per cent of the sample having had the test between ages 10 and 11. Since the Rorschach and TAT were administered by different examiners there was no comparability with respect to inquiry or probing. Thus, the analysis of both the Rorschach and TAT was restricted to the *S*'s spontaneous verbalization to the stimulus before any questions or inquiry were conducted by the examiner. The protocols were scored for the following fantasy categories.

1. *Need achievement on the TAT.* Achievement imagery on the TAT was scored according to the definition of McClelland et al. (8); and themes involving a reference to competition with a standard of excellence were scored achievement imagery.

2. *Rorschach aggression.* The definition of aggressive content on the Rorschach included (*a*) people, animals, or creatures engaged in physical or verbal aggression, e.g., fighting or quarreling, (*b*) explosive objects or explosions, e.g., volcanoes, bombs exploding, fireworks, and (*c*) objects or animal parts normally regarded as instruments of aggression, e.g., spears, rifles, clubs, guns, knives, horns, and claws.

3. *Intellectual curiosity about nature.* For the TAT, curiosity was defined in terms of themes in which someone is interested in the processes or phenomena of nature. Curiosity on the Rorschach was restricted to anatomy or X-ray responses of internal organs or boney parts, e.g., stomach, backbone, ribs.

4. *Passivity.* Because of the limited amount of thematic material in the spontaneous performance, themes of passivity were limited to stories in which the central figure was described as sleepy, tired, or resting.

The fantasy categories were independently scored by the senior author and an assistant without knowledge of the *S*'s IQ scores.[2] Reliability was very high because of the limited amount of content scored for each response and the objectivity of the definitions. Percentage of agreement for the three TAT categories was 95 per cent and for the two Rorschach categories 99 per cent.

RESULTS

Although there was a total of 70 *S*s in the two extreme quartiles, not all of the *S*s had Rorschach or TAT data for the age range under study. Table 1 shows the distribution of *S*s, by sex and direction of IQ change, for the TAT and Rorschach analyses. Because there are approximately twice as many boys as there are girls in Group A, all comparisons were first made separately by sex and results were only combined if the direction of the result for both boys and girls in the same IQ group was in the predicted direction.

1. *Need achievement.* All achievement themes, save one, occurred to Cards 1 and 17 BM. The typical achievement story to Card 1 concerned a boy who wanted to master the violin and/or become a famous violinist, while the typical achievement theme to 17 BM involved competitive activity with regard to rope climbing. Table 2 shows the percentage of *S*s in each group reporting achievement imagery plots to Cards 1, 17 BM, and to both pictures.

TABLE 1

DISTRIBUTION OF *S*s BY SEX AND DIRECTION OF IQ CHANGE USED IN THE ANALYSIS OF THE TAT AND RORSCHACH

Group	TAT		Rorschach	
	Boys	Girls	Boys	Girls
Group A	22	11	22	10
Group D	10	20	9	18
Both groups	32	31	31	28

TABLE 2

PERCENTAGE OF *S*s REPORTING ACHIEVEMENT IMAGERY TO CARDS 1 AND 17 BM

TAT Card	Group A			Group D		
	Boys	Girls	Boys and Girls	Boys	Girls	Boys and Girls
Card 1	36.4	50.0	40.6	27.3	15.0	19.4
Card 17 BM	36.4	30.0	34.4	0.0	15.0	9.7
Cards 1 and 17 BM	22.7	10.0	18.8	0.0	0.0	0.0

For both Cards 1 and 17 BM, more male and female *S*s in Group A report achievement imagery than the boys or girls of Group D. For Card 1, the difference between Group A and Group D girls is reliable at the .03 level; the difference for boys is in the predicted direction but not significant. For Card 17 BM, the difference between Group A and Group D boys is significant ($P = .03$) and in the predicted direction for girls. All *P* values are for one tail and were evaluated using the exact method suggested by Fisher (5). When the sexes were pooled, comparisons between Groups A and D were significant not only for Cards 1 and 17 BM separately but also for the number of *S*s telling achievement imagery to both Cards 1 and 17 BM ($P < .10$, .03, and .01 respectively). Thus, the *S*s who showed increases in IQ were more prone to structure Cards 1 and 17 BM in terms of achievement oriented behavior than the *S*s in Group D.

2. *Aggressive content on Rorschach.* There was no significant difference between Groups A and D or between boys and girls with respect to the mean number of responses per protocol, and the mean for the entire sample was 27 responses. There was no difference between Group A and Group D girls with respect to percentage of each group reporting one or more aggressive responses per protocol (30.0 per cent for Group A versus 33.0 per cent for Group D). However, the difference between Group A and D boys approached significance with 59.1 per

---

[2] The writers wish to thank Mary Schnurer for her assistance in assessing the reliability of the scoring.

cent of the former and 22.2 per cent of the latter reporting one or more aggressive images ($P = .07$). Thus, the prediction of a correlation between IQ increase and aggressive imagery held only for the boys. Because of the tentativeness of this result and the more speculative nature of the hypothesis relating competitive striving and aggressive content, an attempt was made to validate this finding by analyzing a later Rorschach protocol for the boys in Groups A and D. Not all of the boys had Rorschachs administered to them at a later age, and only 15 *S*s in Group A and five in Group D were available for analysis. The median ages at the time of administration were 13-8 and 15-0 for Groups A and D respectively, and there was no significant difference in the lengths of the protocols of the two groups. The results were in the same direction for 86.7 per cent of Group A, and 20.0 per cent of Group D reported one or more aggressive images, and this difference is highly significant ($P = .01$).

3. *Intellectual curiosity*. The only TAT card eliciting curiosity plots was Card 14, and the typical theme described a person gazing at or interested in the stars or the heavens. Table 3 shows the percentage of each group telling such themes to Card 14.

Both the boys and girls in Group A told more themes of interest in the stars or heavens than the males and females in Group D ($P = .14$, $P = .10$, respectively) and combining of the sexes yielded a highly significant difference between Groups A and D ($P < .01$).

4. *Anatomy and X-ray responses on the Rorschach*. There was no difference between Group A and Group D girls reporting one or more anatomy responses (30.0 per cent versus 38.9 per cent for Groups A and D respectively). For the boys, 31.8 per cent of Group A and 0.0 per cent of Group D reported anatomy or X-ray imagery, a difference that approached significance ($P = .06$). This finding was also validated on the same sample of 20 boys that was used to check the differences in aggressive content. The results were in the same direction with 60.0 per cent of Group A and 20.0 per cent of Group D reporting anatomy content ($P = .15$).

TABLE 3

Percentage of *S*s Reporting Themes of Curiosity to Card 14

Sex	Group A	Group D
Boys	40.9	18.2
Girls	30.0	5.0
Boys and girls	37.5	9.7

TABLE 4

Percentage of *S*s Reporting Themes of Passivity to Card 3 BM

Sex	Group A	Group D
Boys	9.1	27.3
Girls	10.0	45.0
Boys and girls	9.4	38.7

5. *Passivity*. Card 3 BM accounted for most of the passivity themes and the groups were compared with respect to the incidence of stories to Card 3 BM in which the central figure was sleepy, tired, or resting. Table 4 shows the percentage of each group telling such themes. Both the boys and girls in Group D showed more passivity themes than the boys and girls in Group A. Although only the difference for the girls was significant ($P = .06$), when the sexes were pooled the difference was highly reliable ($P < .03$).

Cards 3 GF, 5, 6 BM, and 12 F did not furnish data relevant to the hypotheses under test and these results are not summarized.

Discussion

In the main, the hypotheses about the differences between Groups A and D have been verified. Boy and girl ascenders produced more TAT achievement imagery and curiosity about nature than Group D children and male ascenders displayed more aggressive content on the Rorschach than the boys in Group D. The higher incidence of aggressive imagery for the boys who gained in IQ was interpreted as reflecting stronger competitive motivation. Finally, the *S*s in Group D were presumed to have a more passive orientation since they were more likely to perceive the ambiguous figure on Card 3 BM as sleeping or tired. The relation between Rorschach anatomy content and IQ gain was the most tentative finding.

The results are interpreted as indicating that high motivation to achieve, competitive strivings, and curiosity about nature may motivate the acquisition of intellectual skills and knowledge which, in turn, facilitates increases in tested IQ. If one accepts the generally assumed notion that boys are more competitive and achievement oriented than girls, the fact that there were twice as many boys in Group A as there were girls supports the present interpretation. A recent study using the Edwards Personal Preference Schedule found that high school boys obtained higher need achievement scores than high school girls (7).

These results are not interpreted as indicating that strong achievement, competitive, and curiosity motives are the only variables involved in producing gains in IQ. The *S*s in this study are all average or above in IQ and there is not adequate sampling of children with lower IQ levels. One would not expect *S*s with low IQs or language handicaps to suddenly show an interest in reading despite achievement needs or intellectual curiosity. The child who spends increased time reading because of a heightened interest in natural processes must have already learned the basic reading skills so that this behavior is not a difficult or unlikely choice for him.

Similarly, needs for achievement and successful competition should only motivate attempts at improvement of intellectual abilities in a social milieu where praise, recognition, and superior status are awarded for such accomplishment. That is, achievement-oriented children from homes in which intellectual activity was praised would probably be more likely to master intellectual skills than achievement-oriented children from homes in which such accomplishment was not rewarded. In a cultural environment where athletic ability, fighting prowess, or success with the opposite sex was highly valued, one might expect the child to choose these behavioral channels to gratify his achievement and competitive needs. The parents in the Fels population are predominantly middle class and tend to place importance on intellectual accomplishment. A large majority of the parents have attended college, and since enrollment in the Fels program is voluntary it might be inferred that only parents who valued knowledge and scientific pursuits would be predisposed to become part of the research population. Thus, the children under study tend to come from homes which value intellectual ability.

Study of the educational attainment of the parents of the *S*s in Groups A and D revealed no significant difference between the groups with respect to the percentage of families in which both parents attended college (57.1 per cent for Group A versus 42.9 per cent for Group D; $P > .30$). Although there is a slight difference favoring the educational level of Group A families, the difference was not dramatic. There may be important differences between Groups A and D with respect to the differential encouragement of intellectual achievement, but measurement of these differences would probably require variables more refined than educational level of the parents. However, even though parental emphasis on intellectual activity may increase the child's desire to improve his cognitive skills, the child's predisposition to adopt or rebel against parental values should selectively influence his motivation to strive for intellectual accomplishment. Thus, the type of relation between parent and child may be an important factor in this process.

Finally, there is the possibility that genetic and/or constitutional variables may play a role in facilitating marked IQ changes. There is considerable data indicating that genetic factors influence general IQ level but less evidence relevant to the role of these variables in producing childhood increases in IQ score. For most of the children in our population, IQs tend to level off during the ages 6–10 and most of the marked changes in level occur during the preschool years. However, the exact relationship between genetic variables and IQ change has yet to be determined. The phenomenon of IQ increase during the school years is admittedly complex and it is not implied that the child's motives are the major factor. However, it is suggested that personality needs may influence this process. Perhaps the most accurate generalization is that for middle-class children with average or above IQ levels, strong achievement, competitive, and curiosity needs may facilitate IQ gains by motivating the child to master intellectual skills.

A final implication of these findings is that they add indirect evidence for the usefulness of the Rorschach and TAT as research instruments. Validation of a predicted relationship between TAT achievement imagery and IQ gain increases one's confidence in the hypothesis that TAT plots can serve as an index of achievement-oriented tendencies. The results of the Rorschach analysis suggest that aggressive content may be an index of an individual's aggressive predispositions but not necessarily a measure of his tendency to express direct, physical aggression. Although Sontag, Baker, and Nelson (11), using behavioral observations, rated the boys in Group A as more competitive than those in Group D, there was no difference between these groups with respect to intensity or incidence of direct verbal or physical aggression or destruction of property. We have assumed that competition is a socially approved form of aggressive behavior and the higher incidence of aggressive content for Group A boys was presumed to be a result of their more intense competitive strivings. Some clinicians who use projective tests are too prone to focus on predictive statements about direct, physical aggression when confronted with a protocol containing aggressive content. One is apt to overlook the fact that the individual may have alternative behavioral channels for expression of aggressive motives.

SUMMARY

For a group of 140 boys and girls in the Fels Research population on whom continuous Binet IQ data were available, a distribution of IQ change was obtained by subtracting each *S*'s smoothed IQ at age 6 from his smoothed IQ at age 10. This distribution of differences was divided into quartiles, and the Rorschach and TAT protocols of the upper, (maximum increase in IQ) and lower (maximum decrease in IQ) quartiles were analyzed and compared. The results showed that in comparing the *S*s who showed IQ increases with those showing IQ decreases, the former had, on the TAT, significantly more (*a*) achievement imagery on Cards 1 and 17 BM and (*b*) themes of curiosity about nature on Card 14, and significantly fewer themes of passivity on Card 3 BM. For the boys only, more of the *S*s who increased in IQ had anatomy responses and aggressive imagery on the Rorschach. The results were interpreted as indicating that high need achievement, competitive striving, and curiosity about nature are correlated with gains in IQ score because they may facilitate the acquisition of skills that are measured by the intelligence test.

REFERENCES

1. BAKER, C. T., SONTAG, L. W., & NELSON, VIRGINIA L. Specific ability in IQ change. *J. consult. Psychol.*, 1955, **19**, 307–310.
2. BAYLEY, NANCY. Mental growth in young children. *Yearb. Nat. Soc. Stud. Educ.*, 1940, **39**, (II), 11–47.
3. BAYLEY, NANCY. Consistency and variability in the growth in IQ from birth to eighteen years. *J. genet. Psychol.*, 1949, **75**, 165–196.
4. BRADWAY, KATHERINE. IQ constancy on the Revised Stanford-Binet from the preschool to the junior high school level. *J. genet. Psychol.*, 1944, **65**, 197–217.
5. FISHER, R. A. *Statistical methods for research workers.* (5th ed.) Edinburgh: Oliver & Boyd, 1934.
6. HAGGARD, E. A., DAVIS, A., & HAVIGHURST, R. J. Some factors which influence performance of children on intelligence tests. *Amer. Psychol.*, 1948, **3**, 265–266.
7. KLETT, C. J. Performance of high school students on the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1957, **21**, 68–72.
8. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
9. RICHARDS, T. W. Mental test performance as a reflection of the child's current life situation: A methodological study. *Child Develpm.*, 1951, **22**, 221–233.
10. SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Personality as a determinant of performance. *Amer. J. Orthopsychiat.*, 1955, **25**, 555–562.
11. SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Mental growth and personality development. *Monogr. Soc. Res. Child Develpm.*, in press.
12. WELLMANN, BETH L., & MCCANDLESS, B. R. Factors associated with Binet IQ changes of preschool children. *Psychol. Monogr.*, 1946, **60**, No. 2 (Whole No. 278).

Received April 29, 1957.

THE PSYCHOLOGY OF SMOKING[1]

CHARLES McARTHUR, ELLEN WALDRON, AND JOHN DICKINSON
Harvard University

Who smokes? Who doesn't smoke? Who smokes too much? Who can stop smoking? Reviewing the sparse literature on smoking and personality, we could find neither descriptive data answering these questions nor good theory from which the answers might be deduced. Therefore, we have made a first attempt at delineating a psychology of smoking.

Our basic data come from The Study of Adult Development, once known as the Grant Study. That study has been described elsewhere (6). Our subjects were a panel of 252 Harvard alumni who were selected during their sophomore years for lack of visible abnormality. During their sophomore years, which fell between 1938 and 1942, they were studied by a great range of medical, physiological, psychological, anthropological, and sociological techniques. We have been following these men since by annual questionnaire, retesting, and visits. Questions about smoking habits were routinely asked in the questionnaires during the fifteen or so years the men have been followed. This very full material permitted the study staff to make comparisons between smoking habits and hundreds of variables. Only the findings relevant to psychosocial theory are reported here.[2]

WHO DOESN'T SMOKE?

A striking number of our subjects did not smoke. When the men were first studied, as sophomores, 108 of them (some 45 per cent) had not begun to smoke at all. Sixty-one never have. That is about 24 per cent of the group. In any one year, more than 40 per cent of the men will be "off" smoking, though for some this is very temporary.

Our theory begins with Bales' very general propositions about the origins of any compulsive habit. Bales (2) began by studying compulsive drinkers. He found that "analysis shows the craving to be a result of at least two types of underlying elements: (1) some need or complex of needs. . . and (2) an orientating structure of habitual thought patterns and associated emotional justifications." Orientation has to come first; if the experience is found gratifying, it may be put into the service of needs. As Bales (3) says, "it is during the formation of the compulsive habit that the influence of the society and its culture come to a critical focus in the individual person and work their effect." In particular, it is society that determines whether the new habit seems to be "a means of relieving. . . inner tensions, or whether such a thought arouses a strong counteranxiety."

Convergent theories of anthropologists, sociologists, and psychologists, reviewed and empirically confirmed elsewhere (10, 13, 14,15), all suggest that the use of tobacco to resolve major personal needs might "arouse a strong counteranxiety" in upwardly mobile members of America's lower middle class. It seems likely that this group differs in their attitude toward smoking from people who stand both above and below them in the social hierarchy. Their attitude to smoking would, presumably, be corollary of a basic value system that these people hold in contrast to most of contemporary society.

The source of this value system is described by Weber (23) as the Protestant Ethic. It is a work morality, emphasizing the Devil's stake in idleness and self-indulgence. It is an ethic that abhors as Sin the wasting of one's substance in unprofitable trivia, since such waste, in the original protestant dogmas, was evidence that a man had not been vouchsafed Grace. This ethic was, as Mills (18) remarks, a morality of producers, who abhorred consumers. It is typical that, after one especially long

---

[1] This study was supported by the Tobacco Industry Research Committee. It was conducted as part of the research program of Harvard's University Health Service.

[2] Some of the other findings of the study staff, especially those in relation to medical and physical variables, are reported in Heath, C. Differences between Smokers and Non-Smokers, in press. A few items appear in both papers, as they seem relevant.

A table, giving data for each of the findings reported, has been deposited with the American Documentation Institute. Order Document No. 5464, remitting $1.25 for 35 mm. microfilm or $1.25 for 6 by 8 in. photocopies.

Calvinistic debate, tobacco was settled upon as a good thing because the growing of it produced profit.

This theology of "provident purposes" was soon secularized in the form of "worldly asceticism," as Weber illustrates from the diaries of Benjamin Franklin. Indeed, the process of lay morality descending from religious dogma is still going on, even among the Harvard group we are studying, as Allport (1) has shown in his study of religion in the postwar college student. It is thus that the traditional middle-class values were generated.

These values are now being replaced in our society, according to sociologists. They have become the property, though perhaps not exclusively, of the "Old Middle Class." They belong to Riesman's (20) "Inner-Directed" group, so soon to be replaced by the Outer-Directed; to Mills' (18) "cheerful robots," so soon to become equally cheerful mass consumers. Yet it is these values that have generated what Mead (16) calls "the American core culture" and what Kluckhohn (10) describes as the dominant value profile in American culture and expects to find focussed in the middle class, and which McArthur (13, 14) has empirically shown to be focussed in the lower middle. It still gives rise in the American middle class to a new generation instilled with the "achievement mores" whose operation in a contemporary urban setting is described by McClelland (15). Children reared in this tradition are mobile, following what Miller and Form (17) describe as Ambitious Careers. As one study (14) shows, "The Future, Doing-oriented family must produce sons reared in the 'achievement mores,' taught to look forward to by-passing or surpassing their fathers' occupational roles.... It is the hopes of the mother that these sons must realize in order to feel successful. They... will have introjected her precepts. It is these boys who will, after college, be expected to leave the family and 'make their own way.'" Children reared in this tradition are still (15) weaned and trained early, encouraged to be ambitious, hard-working, and clean-minded. "I know that Andy has a clean mind," wrote one father of a nonsmoker, "he does not smoke, he does not loiter on the street with gangs, he has been brought up to look for the fine things in life and only by hard work and perseverance can they be achieved and enjoyed."

In the study, we have found (13) that a simple empirical tool for bringing out middle vs. upper class contrasts is to compare Harvard men who prepared for college in public high schools with Harvard men who prepared for college in private preparatory schools. The high-school graduate who goes to Harvard has usually received a heavy dose of "work morality" during his rearing. As an earlier study (14) suggested, "His role in the family often is that of standard-bearer, the son who is to win status or ethnic mobility." The prep-school boy who goes to Harvard is more likely to have been bred on a code of acceptable behavior. Tradition sets bounds on him in terms of "what is done," not in terms of "worldly asceticism" as a means of "getting ahead." If a concentration of nonsmokers exists in our data, then we may well predict that it will focus among our public-school graduates.

This is the case. In a steady progression with status, 20 per cent of the graduates of very exclusive private schools do not smoke, 30 per cent of the graduates of less exclusive private schools, and 40 per cent of the graduates from public schools. This pattern is significant at the .01 level. Part of the explanation of this fact may be that the prep-school boys learned to smoke in their dormitories (rules notwithstanding) while the high-school graduate did not experience dormitory life with its peer-orientation until coming to college. Some facts fit this notion. In retrospect, about three-quarters of the private-school graduates who smoke say that they learned to smoke in prep school or at the time of prep school; about three-quarters of the public-school graduate smokers say they learned to smoke in college. As one progresses from private boarding to private day to public day schools, there is a significant (.01) decline in the proportion of boys who already smoked at the time they were taken into our study. However, there is a status difference operating together with this dormitory factor. If we hold boarding constant, there remains a significant (.02) difference between the more exclusive and the less exclusive schools. Such a pattern may well have been generated in the families who elect to and are able to send their sons to these schools, rather than in the schools themselves. Besides, the tacit other half of the dormitory argument is that high schools do not orient their pupils toward smoking. This cannot be said. Hollingshead (8) tells us that in "Elmtown" high school, "practically all boys (91 per cent) smoke, irrespective of age or (social) class." Our public-school nonsmokers are therefore somehow special. It may be supposed that their rearing for mobility, in terms of the old ethic described earlier, may have created a considerable "strong counteranxiety" to undermine the orientation from their peer culture toward learning to smoke.

> In Elmtown,
>
> Law and the mores deny high-school students the right to enjoy the pleasures derived from tobacco, gambling, and alcohol. However, the mystery with which adults surround these areas of behavior lends them a special value which seems to act as a stimulus to many young people who desire to experience the supposed thrill of pleasures their elders deny them. The conspiracy of silence which is an essential part of the clandestine violation of the mores has already taught them how easy it is to avoid restrictions imposed by law and taboo if they are discreet about how, where, and under what circumstances it is done. Acquisition of knowledge of the means of transgressing against alcohol, tobacco, and gambling taboos without being caught and the thrill of violating these taboos take place for the most part in the clique.

Presumably any peer-oriented youngster is likely to smoke, whatever his social status, as part of a kind of defiant claim on adult status. It may be supposed that the more earnest students in high school take adult prohibitions more seriously and, perhaps, do not always fit well with "the clique." Theirs may be a more compliant claim on adulthood. Perhaps among the earnest nonsmoking nine per cent of Elmtown seniors, there was one who applied to Harvard. More went to other colleges and were mobile into white collar technical and professional occupations. It is not to be supposed that attendance at any one university is a potent cause. Census data (5) show a nationwide increase of nonsmokers when professional and technical people are compared with other occupational groups. Nonsmoking in America is a social class phenomenon.

It is in England, too. A Hulton survey (7) suggested that nonsmoking was slightly commoner among men of the English middle class while heavy smoking was commoner among English working-class men. An excellent market survey[3] in England not only confirms this difference in rate of smoking but also documents in many ways the fact that nonsmoking is widely perceived as a mark of "middle-class respectability."

Our data certainly suggest that smoking is popularly defined as faintly disreputable or, at least, one of the small vices. Nonsmokers are significantly often (.01) nondrinkers; indeed, it seems that, as Straus and Bacon (22) report from Yale, drinking precedes smoking as the student takes up adult foibles. Our nonsmokers are also significantly often (.01) people who do not drink coffee.

We may expect some orientation toward nonsmoking to have been taught these men by their religions. The kind of piety that Weber describes should spill over into self-denial, even in the absence of a specific theological proscription. Proscription of smoking sometimes does occur. Among Protestant denominations, it is usually those of less social standing (19)—e.g., Fundamentalists—who condemn smoking, while those of more fashionable status —e.g., Episcopalians—do not. The lower-middle-class focus of nonsmoking may be thus overdetermined. At any rate, we may find overlap among smoking habits, social status, and piety.

Earlier in the course of the study, Heath had set up ratings of the devoutness of the participants and also of their families. These ratings were based on considerable information about theological beliefs, church attendance and activities, personal use of prayer, etc. (No attention was paid to vices, great or small, and hence smoking in itself did not enter this rating.) The boys who did not smoke when they entered the study had more devout parents (.05). They were also more likely (.01) to attend church while at Harvard, despite the absence of any chapel requirement. Lifelong nonsmoking is associated even more strongly (.02) with devoutness of parents. It would seem that the individual lastingly introjects his smoking morality! Indeed, the nonsmokers are often rated higher on devoutness than their parents.

This introjected piety seems to be more an individual matter than we expected. In the study data, nonsmoking is the modal pattern for the members of no one religious denomination. Nor did the interaction between piety and status work quite as expected. It is true that, as expected, most Protestant families with low incomes were rated high on devoutness and produced sons who were nonsmokers. Presumably, we see here the introjection of pietistic standards, at least in the form of "worldly asceticism," in the absence of theological proscription.

---

[3] Davis, F. Cigarette Smoking Motivation Study. This is an unpublished monograph prepared by Research Services Limited, to whom we are grateful for access to it.

The nonsmokers may represent survivals not only of the protestant ethic and middle-class morality but also of the related phenomenon Riesman (20) calls "Inner Direction." For example, in answering the Strong Vocational Interest Blank, more sophomore nonsmokers (.03) said they preferred "nights at home" to "nights away from home," and smokers said the opposite. Also, nonsmokers preferred (.02) "belonging to few societies," smokers "belonging to many societies." These trends were especially marked among nonsmokers from public schools, whose earnest individualism is thus reaffirmed, but private-school nonsmokers showed some of the same trend.

A curious sidelight may be interpreted variously: more sophomore smokers (.01) say they would prefer "preparing the advertising for the machine," in the appropriate Strong item, while nonsmokers prefer not to. It is also true (.05) that more smokers responded "Like" to the occupation of Advertiser, nonsmokers responding "Dislike." Perhaps the nonsmoker is so Inner Directed that he does not have his "radar" tuned to the latest fad in mass consumption. He is, perhaps, not yet part of the consumer ethic that is said (18) to be appearing now among the middle class.

A related finding, based on Strong-like items in a recent questionnaire, is that smokers respond "Like" to the occupation Sales Manager (.01) but "Dislike" to the occupation Scientific Research Worker (.01). This attitude is reflected in career choices: the nonsmokers contribute more physical scientists (.01). These facts overlap the status differences already observed; few private-school boys go into science but many go into business. Whether the ethos of the physical scientist overdetermines his nonsmoking over and above the effect of his middle-class origins and upward mobility is not clear from our data. Certainly the two—mobility and science—when combined pull powerfully: of the twelve public-school boys from low-income families who became scientists, ten never smoked and one tried it but stopped.

In summary, the nonsmoker seems to have been oriented by the mores of a particular American subculture. He is often of lower-middle-class origin and himself upwardly mobile. He shows the "worldly asceticism" that has stemmed from the old Protestant ethic. Often he is pious, perhaps more so than his parents. It seems likely that he has reacted to smoking as being one of the "small vices" to which the flesh is heir. He is, at any rate, an Inner-Directed person, introjecting the morals of his youth, perhaps a serious sort, and maybe an introvert. He does not go along with the suggestions for a consumer morality offered by the mass media. He approves scientific rather than business values and may often himself be a scientist or engineer. Just what causes underlie this web of correlations is not clear, but existing theory about the ethos of the Old Middle Class or about the American core culture seems relevant. The standards this man has introjected furnish sufficiently "strong counteranxiety" to prevent his sharing the orientation toward smoking that seems to be common to all the rest of American society both below and above him.

WHO SMOKES HEAVILY?

Whether a man smokes or no seems best explained by his social orientation. Whether he becomes a heavy smoker seems best explained by his personal needs. Bales (2) speaks of the effectiveness, once a man has been oriented toward adopting a habit, of "the degree to which the culture operates to bring about acute needs for adjustment, or inner tensions, in its members. There are many of these: culturally-produced anxiety, guilt, conflict, suppressed aggression, and sexual tensions of various sorts may be taken as examples."

There is no doubt that our very few really heavy smokers, who meet the criterion for Heavy Smoking now customary in medical research—"two packs a day for several years"—have more than their share of "acute needs for adjustment." Most have had marital problems, some quite dramatic. All are given to impulsive acts, some to physical violence, if only in the form of volunteering for dangerous missions. Several are hard-driving, tough competitors. None are usual for our group. As one observer phrased it, "they are men who live in overdrive!" Their stories are told by Heath.[4]

That the need to smoke heavily might be generated by anxiety is, of course, a commonsense hypothesis. Our data warns us to view it with caution. We have not been able to correlate changes in smoking with changes in the tenseness of a man's life situation, in spite of the "clinical hunch" of staff members that this relationship held. Perhaps anxious smoking is episodic, so that our yearly questionnaire is too coarse a measure of it. Perhaps something is to be learned from the answers to a question about the symptoms of stress that our participants felt when under pressure. Half of the smokers said they "smoked more" in these circumstances. The interesting fact, though, is that 70 per cent of the heavy smokers but only 30 per cent of the light smokers said this. Is smoking, then, a suitable tension reducer only after it has become a firmly ingrained habit? One thinks of learning theory: a response must be "high in the habit hierarchy" before it has the "availability" to reduce tension. Or one may think that this little item is a rather pretty confirmation of Bales' contention that only after thorough orientation can a compulsive habit be used to satisfy anxious needs.

The most striking correlation between heavy smoking and personality is found in our inkblot test material. In the early days of the study, Dr. Wells (24, 25) used an abbreviated Rorschach procedure, allowing one minute for response to each of the standard blots, which he called the Timed Rorschach. The resulting scores are not identical with Rorschach scores but seem to be interpretable by similar theory. The Rorschach variable we concentrated on is the Experience Balance, defined as the ratio of the number of human movement responses to the weighted sum of the various kinds of color responses. If both members of this ratio are equal to or less than 2:2, one speaks of a "coartated" Rorschach, by which one means an emotionally narrow, drily factual, excessively controlled performance. If either member of the ratio exceeds 2, one sort or another of emotional expression has taken place, and the label "coartated" no longer applies.

Heavy smokers produce more (.05) coartated records. (Our definition of Heavy Smoking is now loosened to include men with lifetime averages of a pack a day, so as to include a workable number of cases.) What is more, coartated people tend (.01) to increase the amount of their smoking strikingly as the years go by. This is not a universal trend; many of our men reach a smoking pleateau early or vacillate between nonsmoking and moderate smoking. The coartated person is common among that plurality of our people whose smoking curve "snowballs," showing a positive acceleration. The appearance of these curves suggests to the eye that they be interpreted as showing accelerating habituation. Interestingly, such curves often appear among people who started smoking late. (They are not just "catching up," however; they soon surpass the smoking rates of men who started earlier.)

This association between accelerating smoking rates and coartation is strong enough to override the other trends so far reported. A coartated man tends to become a heavy smoker no matter what his social background or how pious his family—or himself. His orientation may show in his being a nonsmoker in college if he came from a nonsmoking type of background. Soon after college, he typically begins to smoke, and the habit accelerates. One is reminded of Bales' (2) comment that, "there is reason to believe that if the inner tensions are sufficiently acute, certain individuals will become compulsively habituated in spite of opposed social attitudes."

Are our coartated people that badly off? There is no very detailed Rorschach theory (9, 21) from which one might interpret their tests. They do seem to indicate a certain lack of emotional resources—or unwillingness to use them. Understandably, these men were rated by the psychiatrists (26) as inarticulate, pragmatic, and bland. They were not necessarily thought to have inner tensions. They were, at any rate, quite hard to get to know. Many were rated "Just-So" by the psychiatrist; that is, they were compulsive fiddlers, deskarrangers, people who allayed their tensions through fussy activities, of which smoking could understandably become one. This very striking finding about coartation remains hard to explain. Nor have we found the means to cross-validate it. Since the college Rorschach pattern predicts lifetime smoking, another long-term study would be needed for ideal cross-validation.

---

[4] Heath, C. Differences between Smokers and Non-Smokers, in press.

Not all heavy smokers are coartated, of course. Among the noncoartated subjects, there is a tendency for the associations between amount of smoking and some of the psychosocial variables previously discussed to be particularly "clean." It is as though coartation washes out the array of emotional variations between persons as well as, perhaps, the variability of the smoking rate. Among these noncoartated subjects, it does appear that the psychosocial variables have more to do with the difference between nonsmoking and moderate smoking (as discussed above) than with heavy smoking.

In the noncoartated group, moderate smoking and heavy smoking are respectively related (.001) to the psychiatrist's rating of Strong Basic Personality and Weak Basic Personality. One presumes that the latter rating is appropriate to the rather uncontrolled men who were described as being among our very heaviest smokers. It is not true that everyone who smoked more than a pack a day over his adult life was poorly integrated. It is true that poorly integrated people were much more commonly found among these heavier smokers. Again, one sees a pattern such as Bales would predict: orientation leading to moderate smoking but "some need or complex of needs for adjustment" leading to excessive habituation.

In summary, then, we may hypothesize that starting to smoke is largely brought about by one's social environment but that reactions to smoking, once it has started, seem to depend in good part on the personal needs that the newly-established habit is able to gratify. Some people seize on the habit compulsively. These people may often be emotionally constricted types for whom there is great gain in a simple "flight into behavior" or they may be restless, active men, for whom smoking is just one more impulsive activity. It would also seem that anxious people can seize on smoking as a tension reducer if they have already, for other reasons, been oriented toward it. In short, the habit, once well available, increases in strength if it serves well the person's emotional economy.

WHO CAN STOP SMOKING?

As every habitual smoker knows to his sorrow, ability to quit or cut down decreases as smoking increases. If we divide our lighter from our heavier smokers at an adult lifetime average of half a pack a day, the contingency relation between "lighter" and "heavier" vs. "can stop" and "can't stop," as shown in recent responses to questionnaires, gives a p less than .001. Even more striking is the march (.01) of the mean numbers of cigarettes smoked per day during adult life. For men who can stop smoking this figure (computed only during their smoking years) is 9, for men who don't try to stop it is 18, and for those who cannot stop it is 20. The sheer amount of tobacco so far consumed is by far the largest difference between the group of men who can stop and the group who, at least so far, cannot.

The variables that were related to becoming a heavy smoker seem also to bear some relation to ability to stop. Thus, coartated Rorschachs were frequent among people who became heavy smokers but it is also true that, with heavy smoking held constant, the presence of a coartated Rorschach seems related to inability to cut down or stop. Some of our scientists have smoked; they could easily stop. (They were not heavy smokers, of course.) Perhaps the scientists show that it is easier to quit when one's compeers do not smoke: none of our five writers can quit. Among our heavier smokers, the psychiatric labels of Sociable, Strong Basic Personality, and Practical characterize mostly people who can stop; Weak Basic Personality, Asocial, Lack of Purpose and Values, Introspective, Ideational, and Inhibited characterize mostly people who cannot. These are all small trends, however, and mostly occur in small numbers of cases. Yet these relationships seem worth mention because of their congruence with general theory. The variables related to smoking compulsively are mostly "need" variables, as the Bales theory would require.

A signal fact is this: ability to stop smoking is directly proportional to the number of months our subjects were fed from their mothers' breast! The means march (.05) as follows: light smokers who can stop were weaned at 8.0 months; heavy smokers who can stop were weaned at 6.8 months; smokers, mostly heavy, who don't try to stop were weaned at 5.0 months; smokers, mostly heavy, who try to stop but cannot were weaned at 4.7 months. We had previously explored for a possible relationship between smoking and weaning or amount smoked and weaning, but it was not until we explored ability to stop smoking that the relationship to breastfeeding became so cleancut.

Many will wish to explain such a finding away. Certainly we would not argue that weaning is the cause of smoking. The argument must proceed through some *tertium quid*. In this we would agree with Linton (12), who stresses that such a crude datum as date of weaning or any other data "focussed primarily on actual technical operations, without a correspondingly detailed study of the maternal attitudes which accompanied these performances," is naive. The events of infant training are part of a broader pattern and symptomatic of it. "We must remember that these influences operate on the child from birth, and continue to operate on him for a long period of time . . . ." The congruence of infant training, childhood rearing, and adult values is shown by McClelland (15) for two contemporary American subcultures. So, in our data, late weaning was associated with those personality traits that are also related to ability to stop smoking. We do not have to postulate that infantile frustration and adult cigarette smoking are unmediated cause and effect.

Yet, it is a commonplace that people who stop smoking seek other oral gratifications. All the "small vices" we have correlated with smoking are oral vices: alcohol, coffee, and tobacco are taken in through the mouth. So is sugar, which Heath shows to be related to smoking. More directly to the point: Brozek (4) has experimentally documented the notion that men who give up smoking gain weight.

Theoretical reasons suggest that smoking should be correlated to psychoanalytic "orality." Our empirical data suggests that it is. Groups with the ability to stop smoking contain a smaller percentage of bottle babies. Thumbsucking was more commonly reported (close to .05) for men who continued to smoke. If one does not think of a psychoanalytic explanation, then a drive-reduction theory like that of Levy (11) seems to fit the data. Nor would these ideas be inconsistent with the Bales model: what they suggest is that these "deeper" needs have little or no effect on whether one smokes but great effect on how tenacious the habit, once adopted, may become.

In summary, the ability to stop smoking is grossly related to the amount of tobacco one has consumed. Good mental health, such that one has control over one's habits in general, seems to be relevant. So, apparently, is oral gratification received as an infant. This "orality" factor may be mediated in many ways, but its meaningfulness should not be overlooked.

THE RELIABILITY OF THE FINDINGS

To check the reliability of some of these findings, we drew five per cent random samples of the class of 1958 and the class of 1961 from the files of the University Health Services. (*N*s were 55 and 58.) Every freshman fills in a medical questionnaire that contains an item about smoking as well as several items relevant to the variables related to smoking in the Study of Adult Development. Clearly, only the criterion of smoking in college (or near the start of college) was available. However, insofar as these checks bore out our findings, they showed them to be reproducible over a period (1939 to 1957) of almost twenty years.

Both samples showed an excess (.01) of public-school nonsmokers. In the first sample, the only one for which income data was available, it was noted again that income was not as good a predictor of smoking behavior as was type of preparatory school.

The relationship between seriousness of purpose and nonsmoking may be adumbrated by one datum. Within the public-school group, there is in both samples an excess of nonsmokers among those boys who characterize as Definite their pre-college choice of career. Similarly, in the sample for which income data is available, there is a strong tendency for the boys who are both Protestant and poor not to smoke. These would seem to be likely to be the lads for whom Harvard meant hard work and mobility. Or, viewed another way, these would seem likely to include most of our members of Fundamentalist denominations.

In both freshmen samples, those boys who announce their intention of going into engineering or science tend not to smoke.

The examing physician rated these freshmen on "Personality Integration." For both samples, the percentage of smokers marches upwards as the ratings become less favorable. For the combined samples, the percentages of smokers are: for men rated "A," 0 per cent; for men rated "B," 23 per cent; for men rated "C," 37 per cent; and for men rated "D," 50 per cent. There were few A's and few D's. The contingency table made by combining A and B ratings in one column and combining C and D ratings in the other and setting these ratings against smoking and nonsmoking gives a chi square of 3.6, so p is not quite as low as .05.

A similar rating by the examining physician was a prediction of College Adjustment. The percentages again march (though most clearly for the combined samples) as follows: among those rated "A," there were 13 per cent smokers; among the "Bs," 23 per cent; and among both "Cs" and "Ds," there were 50 per cent. Combining A and B and C and D, one gets a contingency table from which chi square is 8.0 and p is less than .01. Both these findings from the physicians' ratings would seem to parallel the study findings that smoking went with psychiatric ratings suggesting poorer mental health.

The study finding that smoking and drinking are correlated was amply confirmed for the class of '58 but not so strongly shown in '61. In both samples, the tendency for drinking to precede smoking was noted.

On the whole, then, where we could cross-check the study findings, they seemed to hold up well. In this connection, it may be worth pointing out that a few of the patterns reported from the Harvard data have been found elsewhere. An unpublished survey of Yale seniors shows the public-private differences in smoking habits still pronounced at the end of the college career. The priority of drinking to smoking is noticeable. (This was, of course, first noted at Yale by Straus and Bacon.) The introjection of family standards shows very clearly as a source of smoking mores. We have already cited notable corroboration from market research done in England, where the association of nonsmoking with white-collar, middle-class identity was amply confirmed, as well as the definition of smoking as one of the minor vices. The same pattern is suggested by a national sample questioned about their smoking habits by the U. S. Bureau of Census (5). Results already cited from the University of Minnesota are consistent with our findings.

The one area in which no cross-validation has been possible has been that of psychodynamics, especially the findings from the Timed Rorschach. Since these personality patterns are very relevant to a theory of smoking, it is to be hoped that something may be done to check them.

SUMMARY

A large part of what we have learned about the correlates of smoking habits may be made to fit a conceptual model like that of Bales. The fact that a man smokes or does not seems to be determined by whether or no he has been oriented to the habit as a result of his social milieu. Whether he becomes a heavy smoker or is unable to stop smoking seems determined by the usefulness of the smoking habit to his personal needs.

Nonsmokers tend to be lower-middle class in origin, upwardly mobile, earnest young men, bred in a work morality that is conducive to Inner Direction. Their parents and they themselves are often pious. They may pursue scientific or technical careers in many instances. Smokers are (in our data) likely to come from more privileged backgrounds, often entering business or humanistic careers, often having been raised in a Being or Being-in-Becoming orientation. Both subculture and the family as the mediator of the subculture are important determinants of whether and when the young man is oriented to the smoking habit.

Whether a smoker becomes a heavy smoker seems to depend on whether the habit serves many of his important needs. Very anxious or agitated men may adopt smoking as a tension reducer, but this use of smoking seems to be common only when the habit is already well established by other circumstances. Emotionally-constricted individuals seem to "take to" smoking with special eagerness.

Whether a man can alter his smoking habits seems to be most "deeply" determined. Efforts to quit or cut down seem to be normal,

since smoking is quite widely seen as a "small vice," even by smokers. Whether a man can succeed in these efforts is first of all a function of how much tobacco he has consumed in his lifetime. Certain personality variables are also relevant. In general, it is the needs that lead to heavy smoking after the habit has begun that also militate against quitting or cutting down. The social variables that were related to starting to smoke play little role in inability to quit. The effects of oral gratification at the breast may be relevant to control excessive smoking, once a great deal of tobacco has been consumed.

Where data were available, the study findings were cross-validated against samples from two Harvard classes, with generally encouraging results. Some confirmation of other findings has been had from unpublished studies at distant points and from a nationwide study by the Census Bureau.

Not all that we know about smoking fits into the proferred conceptual scheme. However, the scheme outlined makes our major findings hang together and may serve as a first hypothesis in further exploration of the psychology of smoking habits.

REFERENCES

1. Allport, G. The religion of the post-war college student. *J. Psychol.*, 1948, **25**, 3–33.
2. Bales, R. F. Social therapy for a social disorder—compulsive drinking. *J. soc. Issues*, 1945, **23**, 1–9.
3. Bales, R. F. Cultural differences in rates of alcoholism. *Quart. J. Stud. on Alcohol*, 1946, **6**, 480–499.
4. Brozek, J., & Keys, A. Changes of body weight in normal men who stop smoking cigarettes. *Science*, 1957, **125**, 1203.
5. Haenszel, W., Shimkin, M., & Miller, H. Tobacco smoking patterns in the United States. Washington: U. S. Govt. Printing Office, 1956.
6. Heath, C. W. *What people are.* Cambridge, Mass.: Harvard Univer. Press, 1945, p. 141.
7. Hobsen, J., & Henry, H. (Eds.). Patterns of smoking habits. *Hulton research studies of the British social pattern.* London: Hulton Press, 1948.
8. Hollingshead, A. *Elmtown's youth.* New York: Wiley, 1949.
9. Klopfer, B., & Kelly, D. *The Rorschach technique.* Yonkers-on-Hudson: World Book, 1942.
10. Kluckhohn, Florence. Dominant and substitutive profiles of cultural orientations: their significance for the analysis of social stratification. *Soc. Forces*, 1950, **28**, 376–393.
11. Levy, D. Thumb or finger sucking from the psychiatric angle. *Child Develpm.*, 1937, **8**, 99–101.
12. Linton, R. *Culture and mental disorders.* Springfield, Ill.: Charles C Thomas, 1956.
13. McArthur, C. Personalities of public and private school boys. *Harvard educ. Rev.*, 1954, **24** (4), 256–262.
14. McArthur, C. Personality differences between middle and upper classes. *J. abnorm. soc. Psychol.*, 1955, **50** (2), 247–254.
15. McClelland, D., Rindlisbacher, A., & de Charms, R. Religious and other sources of parental attitudes toward independence training. In D. McClelland (Ed.), *Studies in motivation.* N. Y.: Appleton-Century-Crofts, 1955. Pp. 389–390.
16. Mead, Margaret. *Male and female.* New York: Mentor Books, 1955.
17. Miller, D. C., & Form, W. H. *Industrial sociology.* New York: Harper, 1951.
18. Mills, C. W. *White collar.* New York: Oxford Univer. Press, 1951.
19. Niebuhr, H. R. *The social sources of denominationalism.* New York: Meridian Books, 1957.
20. Riesman, D. *The lonely crowd.* New Haven: Yale Univer. Press, 1950.
21. Rorschach, H. *Psychodiagnostics.* Bern: Hans Huber, 1942.
22. Straus, R., & Bacon, S. *Drinking in college.* New Haven: Yale Univer. Press, 1953.
23. Weber, M. *The Protestant ethic.* (Trans. by Talcott Parsons) New York: Scribners, 1930.
24. Wells, F. Mental factors in adjustment to higher education. *J. consult. Psychol.*, 1945, **9**, 51–56.
25. Wells, F. Clinical psychometrics in higher education. *J. nat. Ass. Deans Women*, 1946, **9**, 51–56.
26. Wells, F., & Woods, W. Outstanding traits: In a selected college group, with some reference to career interests and war records. *Genet. Psychol. Monogr.*, 1946, **33**, 127–249.

Received September 27, 1957. Prior publication.

CRITIQUE AND NOTES

ATTITUDE CHANGE AND JUSTIFICATION FOR COMPLIANCE[1]

ARTHUR R. COHEN, JACK W. BREHM, AND WILLIAM H. FLEMING
Yale University

It has been demonstrated in a number of experiments (1, 3, 5, 6) that forcing a person to take an opinion position discrepant from what he privately believes results in change of private opinion toward the enforced one. The question then arises as to what variables affect the extent of private change under such conditions of compliance. This problem has recently been discussed by Festinger (2). According to his analysis, expressing an opinion discrepant from one's privately held position creates "dissonance" or psychological tension having drive character. The tension may, in this case, be reduced by changing one's private position so that it more nearly coincides with the position expressed.

While the direction of the private attitude change may thus be understood, what, according to dissonance theory, controls the magnitude of change? Both the number and importance of reasons in favor of and against expressing the discrepant opinion control the magnitude of dissonance. Given the expression of the discrepant opinion and holding constant the number of reasons for not expressing it, the greater the number and/or importance of reasons for expressing it, the less will be the dissonance. From these considerations we may state the major hypothesis: *with the discrepant expression and reasons for not making it held constant, the greater is the justification for making it, the less will be the magnitude of dissonance and consequent attitude change.*

In addition, with the reasons for and against taking the discrepant position held constant, the stronger the expression of the discrepant opinion, the greater will be the dissonance created and therefore, the greater will be the resultant attitude change.

An experiment by Kelman (4) has examined the effect on private attitudes of using specific nonsocial rewards to force compliance. The present study emphasizes another type of motivational induction, that of giving reasons unrelated to the attitude issue for expressing the discrepant opinion. In the present experiment, then, justification refers not to the attitude issue but to reasons for taking a view discrepant from one's private opinion.

METHOD

Subjects (*S*s) were 92 Yale College undergraduates who were approached in their classrooms and given an anonymous questionnaire. The questionnaire asked for their opinions on a salient issue, instructed them to write an essay *against* their position, obtained their opinions on the issue again, and measured their perceptions of the experimental situation.

The experiment was conducted shortly after a rumor that Yale would soon become coeducational had created considerable stir on the campus. Thus the issue of attitude toward coeducation at Yale was chosen as the salient attitude dimension. It was expected that most students would show an initial bias against coeducation.

The *S*s were first asked to indicate their attitudes toward admitting women undergraduates to Yale. The scale used was a "pick-two" device; from six choice points a ten-point scale was generated. *S*s were then asked to write an essay taking a stand *opposite* to the one initially indicated. This procedure was assumed to constitute the expression of an opinion position discrepant from the *S*s' private opinion. The instructions were the same for both experimental conditions except that the high justification group was given a number of reasons why the *S* should comply with the request to write the essay against his position.

In the *high justification* condition, the *S*s ($N=29$) were told the following: *E* appreciated their help since the study represented considerable research effort on his part; it was a seminar research paper which *E* had to get done; *E* had received a great deal of cooperation, interest, and willingness to help from other students; the ability to stand back and consider the other side of a complex issue was a real asset; considering the other side of the issue was only fair; taking the other side facilitated a good grasp of a complex issue; and the results of such a survey could be of interest to social scientists and the university authorities. They were given these cumulated arguments and urged to do the best possible job. *S*s in the *low justification* condition ($N=63$) were simply given the instructions common to both experimental groups, which asked them to write the most productive, creative, and imaginative essays possible, unequivocally against their own position. In order to minimize the pressure exerted to comply because of the very nature of the classroom situations, the *S*s were assured that they need not comply at all if they did not want to. Such perceived pressure due to classroom constraints might be expected to reduce dissonance and prevent variations in attitude change from emerging.

After writing their essays, the *S*s were again asked to indicate their attitude toward the issue in the light of additional thought, and also to respond to a postexperimental questionnaire which attempted to measure the success of the experimental inductions.

---

[1] This report was prepared as part of the Yale Communications Research Program.

RESULTS

Check on the Experimental Inductions

Data from the postexperimental questionnaire indicated that the experimental inductions were effective in creating differences between the two groups in justification for compliance. The low justification group perceived that less pressure to comply was exerted upon them than did the highs ($p < .02$), they were more self-motivated to write the essay than the highs ($p < .02$), and they showed fewer interfering responses than the highs ($p < .01$).[2]

Amount of Compliance and Attitude Change

All *S*s in both conditions complied. It may thus be inferred that the general force to comply was very compelling, and little attitude change is therefore to be expected. The average change in the high justification condition is .17 and that in the low is .46; both values are in the expected direction but neither is significantly different from no change.

Amount of Justification and Attitude Change

The major hypothesis concerns the relationship between dissonance created by compliance and attitude change toward the assumed position. It was expected that the condition of low justification would create more dissonance and consequently more attitude change than would that of high justification. The obtained difference between the two conditions in mean change, .29, is in the predicted direction but is not statistically significant.

When the mean "before" opinions on the issue were examined for the high and low justification groups, the low group was found to be significantly more negative ($M = 6.50$ for the highs, 7.75 for the lows, $p < .02$). It was expected that negative attitudes toward coeducation would be shown more by upperclassmen who had internalized the Yale values than by lowerclassmen who had more recently come from secondary schools. In line with this expectation, the low justification group was found to contain far more upperclassmen than the high justification group, which contained relatively more lowerclassmen. The p value is beyond the .001 level.

The implication of these relationships between experimental condition, extremity of attitude, and class at Yale becomes evident when the data on extremity of position and attitude change are examined. The results show that extremity of position on the attitude scale is inversely related to change in attitude ($p < .001$). Since more of the extremes are in the low justification group, probably due to the unequal distribution of upperclassmen in the two groups, and since extremity of position inhibits attitude change, it is difficult to test the hypothesis put forth. Therefore, it seems reasonable to eliminate the extremes (all those *S*s who began at 10 plus the few who began at 1) from the sample in order to correct for this bias; the hypothesis thus is tested within a selected population ($N = 28$ for the low justification condition, $N = 17$ for the high justification condition). After selection, the means on the "before" opinion are now virtually identical: 5.80 for the highs, 5.72 for the lows.

One indication that the selection procedure is justified concerns the quality of the essays written. When the selected population is compared with the extremes on the number of original arguments given in their essays, 49 per cent of the selected population give one or more original arguments, while only 28 per cent of the extremes give such arguments ($p < .05$). The extremes thus appear to have been less involved in complying than the selected *S*s; eliminating such differential involvement provides a better test of the hypothesis.

Comparison of the two selected groups on the effectiveness of the experimental inductions shows that they too differ on all three measures: the lows perceived less pressure, were more self-motivated, and had fewer interfering responses (all p values beyond the .04 level).

When the major hypothesis is tested with the selected population, the lows show a mean change of .83, the highs a mean change of .22. The difference in amount of change is significant by one-tailed t test at the .07 level. A tendency is therefore suggested by the data for greater change to occur in the condition where greater dissonance was postulated.

Strength of Induced Behavior

Another assumed determinant of dissonance concerns the behavior which produces dissonance. It should hold that the better the quality of the essays written by the *S*s against their original position, the more dissonance is created and, therefore, the greater should be the attitude change in line with the induced position. The essays were coded for number of original arguments, on the assumption that the improvisation of original arguments is a clear indication of the degree to which the *S* was actually involved in expressing the discrepant opinion. As such, the measure of original arguments represents the strength with which the discrepant opinion was expressed (3, 4). Among the selected *S*s, a relationship between number of original arguments and attitude change was found only for the high justification group, the group where it was assumed little dissonance would be created. This relationship is significant at beyond the .001 level by chi-square test. For the low justification group where high dissonance and high attitude change occur, no additional dissonance is created by high quality of essay and, therefore, there are no variations in attitude change as a function of number of original arguments. In effect, where dissonance is not created by the experimental variables, it can occur as a function of individual variation in the strength of the induced behavior.

---

[2] Except where specifically mentioned, all statistical tests in this report are two-tailed tests.

It might be argued that those who write high quality essays do so because they feel greater justification. This hypothesis is counter to our present reasoning since in the high justification condition, quality of essay is related to attitude change. However, no reliable relationships were found between quality of essay and perceived pressure (experimental checks on justification manipulations). Furthermore, it is clear that this hypothesis will not explain the major finding, namely, that attitude change tended to be greater where manipulated (and perceived) justification was least.

SUMMARY

This experiment has tentatively illustrated some of the processes involved in creating and reducing psychological dissonance and the implications of such dissonance for attitude change in compliance situations. It was hypothesized that increased justification for taking an opinion position discrepant from one's private opinion would lead to decreased dissonance and therefore decreased attitude change toward the discrepant position. *S*s were asked to write an essay taking a stand opposite to their initial opinion on a given issue. One group of *S*s was then given a number of reasons for compliance with the request (high justification condition). For another group (low justification condition) *S*s were given no detailed justification for writing the essay. Attitude change was determined from a postexperimental questionnaire.

Three sources of dissonance were explored: (*a*) for the whole population, low dissonance and thus little attitude change were obtained because of the forceful nature of the experimental situation which brought general compliance; (*b*) high dissonance and therefore high attitude change tended to be associated with the low justification condition; and (*c*) in the high justification condition, variations in the strength of the induced behavior brought variations in dissonance and thus subsequent variations in attitude change toward the assumed position. These results should be accepted with caution since they arise from *post hoc* analyses.

REFERENCES

1. BURDICK, H. The compliant behavior of deviates under conditions of threat. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.
2. FESTINGER, L. *Theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.
3. JANIS, I. L., & KING, B. T. The influence of role playing in opinion change. *J. abnorm. and soc. Psychol.*, 1954, **49**, 211–218.
4. KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, **6**, 185–214.
5. MCBRIDE, D. The effects of public and private changes of opinion on intra-group communication. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.
6. RAVEN, B. H. The effect of group pressures on opinion, perception and communication. Unpublished doctoral dissertation, Univer. of Michigan, 1953.

Received March 27, 1957.

ANALITY, POLITICAL AGGRESSION, ACQUIESCENCE, AND QUESTIONNAIRE CONSTRUCTION

MAURICE L. FARBER
University of Connecticut

IN a recent paper (3) Rabinowitz attempts to demonstrate that Farber's (2) reported correlation between anality and political aggression can be explained away as a result of an acquiescence response set. There appear, however, to be serious defects in Rabinowitz's methods that render his conclusions doubtful. Christie, Havel, and Seidenberg's (1) recent cogent and detailed analysis of acquiescence is directly relevant; the present note will therefore be limited to brief comment on several features specific to Rabinowitz's study.

1. Rabinowitz's critique is based upon the construction of new Anality and Political Aggression scales which are scored negatively for these characteristics. In this difficult task (*vide* Christie *et al.*) he does not seem to have succeeded. Two items of

his five-item Anality scale (the second and fifth) and three of his four-item Political Aggression scale (first, second and fourth) appear highly doubtful. For example, is a "no" answer to "When I go to a restaurant I like to order foods which I have never tasted before" an anal reply? Or, is disagreement with the statement "The foreign policy of the United States should be based on the assumption that most other countries are our friends, not our enemies" an indication of political aggression? The rejection of a simplistic, soft-headed liberalism is not identical with active political hostility. Farber's anality items hew very closely to the Freudian formulations, and his political aggression items involve affirmation of actively aggressive feelings.

It is therefore not surprising that the correlation between Rabinowitz's scales and Farber's is but +.12 for anality and +.16 for political aggression. While no ultimate measure of validity is available for either set of scales, it is clear that the identically labeled scales are measuring something quite different. Since Rabinowitz's method requires psychologically equivalent scales that are worded in opposite directions, his failure to achieve these renders his critique irrelevant.

2. No item analysis is presented of Rabinowitz's scales, nor, perhaps more important, is there any indication that the discriminatory power of his items was tested. In view of the lack of theoretical relevance of some of his items as well as their possibly low discriminatory power, these omissions are serious ones.

3. Rabinowitz chose to drop Farber's modified Likert-type of scoring of the Political Aggression scale and to substitute simple agree-disagree categories. By so doing he precluded the possibility of isolating and properly weighting the genuinely aggressive response expressed in the "agree strongly" category, thereby dissipating much of the value of the scale. There is evidence, for example, that acquiescence responses are most frequent among those who agree slightly but relatively rare among those who agree strongly (1).

When one considers these methodological deficiencies in conjunction with Christie's demonstration of the small role played by acquiescence in items similar to those employed here, the lack of validity of Rabinowitz's strictures upon Farber's reported relationship is apparent.

REFERENCES

1. CHRISTIE, R., HAVEL, J., & SEIDENBERG, B. Is the F Scale irreversible? *J. abnorm. soc. Psychol*, 1958, **56,** 143–158.
2. FARBER, M. L. The anal character and political aggression. *J. abnorm. soc. Psychol.*, 1955, **51,** 486–489.
3. RABINOWITZ, W. Anality, aggression and acquiescence. *J. abnorm. soc. Psychol.*, 1957, **54,** 140–142.

Received April 10, 1957.

INDEPENDENCE AND RESISTANCE TO PERCEPTUAL FIELD FORCES[1]

DOUGLAS N. JACKSON
Pennsylvania State University

THIS study was designed to test an hypothesis concerning consistencies in cognitive and personality organization. A second purpose was to provide further validation for previously observed relationships reflecting these hypothetical linkages. The general hypothesis may be stated as follows: individuals are consistent in their styles of response to field forces at different conceptual levels. Thus, *S*s who show personality characteristics consistent with independence from group pressure should also be able to resist forces in the perceptual field.

Jackson (4) has developed a reliable measure of the ability to resist reversals of perspective and has obtained a significant relationship between this measure and the tendency to disagree with F-scale items. Hypotheses provided by Witkin (7) and by Linton (6) regarding consistencies in perception and personality led the author to view this task as one in which *S* must cope actively with the perceptual field forces that cause figure and ground to reverse. Without necessarily invoking the hypothetical "brain fields" of the Gestaltist, it is possible to consider the inevitable and lawfully occurring reversals of perspective as manifestations of hypothetical forces in the visual field. Individual differences on this measure might be considered more generally as an indication of an *S*'s level of energy or activity in organizing perceptual and cognitive field forces, rather than submitting to them. In testing this hypothesis, Jackson (5) found in a preliminary study that two criterion measures of this perceptual dimension correlated significantly and negatively with a measure of behavioral conformity and one of passivity on a group of males. The present study seeks to explore further the generality of this hypothetical dimension with a more economical group procedure.

---

[1] This research was completed while the author was at The Menninger Foundation as a U. S. Public Health Service Clinical Research Fellow of the National Institute of Mental Health. The author wishes to thank Drs. N. H. Pronko, D. T. Herman, and G. E. Rice of the University of Wichita for making their students available as *S*s for this experiment. Dr. Riley Gardner's constructive comments concerning the manuscript are gratefully acknowledged. Portions of this paper were read at the Amer. Psychol. Ass., Chicago, 1956.

METHOD

To order people conveniently with respect to personality characteristics associated with independence of judgment, the personality items that empirically distinguished "independents" from "yielders" in conformity experiments by Barron (1) and by Crutchfield (2) were combined into an "inventory." Barron, using Asch's procedure, and Crutchfield, using a similar technique, placed *S* in a situation where experimentally created group pressure to conform was put in conflict with veridical perception and judgment. *S* had to choose between violating his own perception or publicly facing the prospect of being at variance with the judgment of the group. Behavior in these situations provides an index of an *S*'s susceptibility to external influence, of his capacity to withstand social field forces.

The inventory contained 45 true-false items. It was administered to groups of college students, 89 men and 54 women in all. The age range was 17–40, with a median of 20 years.

Two groups, "independents" and "yielders," were selected from the extremes of the distribution of "conformity inventory" scores. The *S*s were administered a group procedure adapted from one developed by Hochberg and McAlister (3) for determining the time one phase of a reversible figure appears to be in perspective. In this technique, a large India ink drawing of a Necker cube (11⅝ in. wide) was presented with explanation of the nature of the reversals of perspective. After a brief practice trial, *S*s were asked to attempt to hold in perspective the "up" phase (which was outlined) and to indicate by writing a "D" or a "U" which phase was in perspective just prior to the sounding of pure-tone, half-second "beeps." They were cautioned to keep their eyes on the cube throughout the experimental series. To assure an equivalent procedure for all groups, the instructions and the pure tone signals were recorded on tape and played through loudspeakers. There were two trials of two minutes each. Each trial contained 24 of the randomly spaced signals. The signals were randomized to prevent them from being synchronous with rhythmic reversals that occur in some *S*s. The score representing susceptibility to the visual field forces in this situation was the total number of "D" responses in the two two-minute trials.

RESULTS

To test the hypothesis that "independents" would manifest significantly greater activity in coping with the visual field forces in the reversible figure, scores from the "conformity inventory" were tabulated and ranked. Because there was a small but significant difference between males and females on the inventory scores, the 27 per cent extreme groups were selected separately for each sex, thus, in effect, partialling out the influence of sex in making the major comparison. There was a total of 36 "independents" and 36 "yielders" selected from the extremes of the distribution of 143 *S*s. These extreme groups were then compared with respect to the perceptual scores, the number of "down" responses in the two trials of two minutes each. The differences are in the predicted direction; the "independents" showed a mean of 11.4 responses indicating that the "down" phase was in perspective (they were instructed to maintain the "up" phase) while the "yielders" had a mean of 17.5 "down" responses. A single classification analysis of variance between the two extreme groups indicated that this difference was significant ($F = 17.56$; $df = 70$; $p < .001$). Thus, evidence has been obtained for the hypothesis that "independents" may be more active in manipulating perceptual field forces.

Some interesting sidelights were obtained concerning some properties of the particular measures used in the study. For example, the corrected split-half reliability of the group measure of perception (.84) was found to compare fairly well with that obtained under conditions of individual administration, where it is usually above .90. The odd-even reliability of the conformity inventory, on the other hand, was quite low; after correction for length it was only .54 for the men and .44 for the women. Lack of homogeneity is often a serious problem in empirically derived indices such as the present one. It argues against the use of this scale for measuring attributes of particular individuals, although it may be of some use in discriminating broad groups for research purposes. On the inventory, females were found to show a slight but significant tendency ($p < .01$) to score in the "conforming" direction as compared with males. This finding is consistent with the type of developmental theory discussed by Witkin and his associates (7) and by others. Significant sex differences have not been encountered in the present measure of reversals of perspective, however.

It is interesting to note that 31 of the 45 conformity items must be answered *true* to be scored in the conforming direction. Acquiescent response-set may therefore be an important aspect of this measure. Such an interpretation is consistent conceptually with the independence-conformity dimension. Jackson's earlier finding (4) of a significant relationship between the perceptual measure and the F scale also may be consistent with this interpretation.

The evidence suggests that those who characteristically resist certain hypothetical perceptual field forces also have the personality characteristics associated with independence of judgment, a form of resistance to social field forces. Those *S*s who tended to answer personality items in the independent direction appeared to show greater activity in resisting reversals of perspective.

REFERENCES

1. BARRON, F. Some personality correlates of independence of judgment. *J. Pers.*, 1953, **21**, 287–297.
2. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
3. HOCHBERG, J., & MCALISTER, E. Quantitative approach to figural "goodness." *J. exp. Psychol.*, 1953, **46**, 361–364.
4. JACKSON, D. N. Cognitive energy level, acquiescence, and authoritarianism. *J. soc. Psychol.*, in press.
5. JACKSON, D. N. Stability in resistance to field forces. Unpublished doctoral dissertation, Purdue Univer., 1955.
6. LINTON, HARRIET B. Dependence on external influence: correlates in perception, attitudes, and judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 502–507.
7. WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

Received April 23, 1957.

EXTRAPUNITIVENESS AND ETHNIC ATTITUDE

GERALD S. LESSER

Hunter College

ACKERMAN and Jahoda explain the complete absence of deep depression in their anti-Semitic psychoanalytic patients by stating, "The existence of an anti-Semitic reaction presupposes a tendency to blame the outside world rather than one's self, and dynamically, such a tendency is in contradiction to the overtly self-destructive trend of a genuine depressive" (1, p. 243). However, other studies (e.g., 3) have reported inconclusive results concerning the relationship between extrapunitiveness and prejudice. The present study examines the relationship between anti-Semitism and extrapunitiveness among both Jewish and non-Jewish boys.

METHOD

The subjects were 20 white non-Jewish boys (ages 10-0 to 13-1) and 24 white Jewish boys (ages 10-2 to 13-2) drawn from one fifth-grade and two sixth-grade classes in two public schools in New Haven, Connecticut. The schools are in adjacent districts and the families constitute a relatively homogeneous upper lower-class group.

Extrapunitiveness was measured through the Children's Form of the Rosenzweig Picture-Frustration Study (5). The mean and standard deviation of the percentage of extrapunitive responses for the non-Jewish boys were 54.90 and 15.04 respectively, and for the Jewish boys were 55.24 and 13.62, respectively. These means are not significantly different.

Modified sociometric procedures have been frequently employed as a method for the indirect assessment of social and ethnic attitudes in children (e.g., 2, 4). Although variables other than prejudice also affect friendship selection, these sociometric techniques applied to the measurement of ethnic attitudes assume that the degree to which a child accepts or rejects the friendship of members of an ethnic group is related to the degree of prejudice he directs toward that group. In the present study, each child was asked to select his three best friends in his class and also to select the three boys that he wished were not in his class at all. An "anti-Semitism" score was obtained for each boy by subtracting the number of Jewish boys he chose for friends from the number of Jewish boys he rejected. A range of scores of +3 to −3 was obtained; the mean and standard deviation for non-Jewish boys were 0.00 and 1.42 respectively, and for the Jewish boys were −0.44 and 1.36, respectively. These means are not significantly different.

RESULTS

Two Spearman rank-difference correlation coefficients (rho) were computed. For non-Jewish boys ($N = 20$), the correlation between extrapunitiveness and "anti-Semitism" is +.60 ($p < .01$). For Jewish boys ($N = 24$), the corresponding correlation is +.48 ($p < .01$). These coefficients are not significantly different.

CONCLUSIONS

Previous studies (e.g., 3) have not found clear evidence of a relationship between extrapunitiveness and prejudice. The present study reports significant positive relationships between extrapunitiveness and anti-Semitism among both Jewish and non-Jewish boys.

Sarnoff (6) obtained a strong negative relationship between extrapunitiveness and anti-Semitism for a group of Jewish adults. For Jewish pre-adolescent boys, the present study found the reverse relationship, i.e., a significant positive relationship between extrapunitiveness and anti-Semitism.

REFERENCES

1. ACKERMAN, N. W., & JAHODA, MARIE. The dynamic basis of anti-Semitic attitudes. *Psychoanal. Quart.*, 1948, **17**, 240–260.
2. CRISWELL, JOAN H. A sociometric study of race cleavage in the classroom. *Arch. Psychol.*, 1939, No. 235.
3. LINDZEY, G. An experimental examination of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1950, **45**, 296–309.
4. MUSSEN, P. H. The reliability and validity of the Horowitz Faces Test. *J. abnorm. soc. Psychol.*, 1950, **45**, 504–506.
5. ROSENZWEIG, S., FLEMING, EDITH E., & ROSENZWEIG, LOUISE. The children's form of the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1948, **26**, 141–191.
6. SARNOFF, I. Identification with the aggressor: some personality correlates of anti-Semitism among Jews. *J. Pers.*, 1951, **20**, 199–218.

Received June 16, 1957.

ANTICIPATED AUDIENCES AS DETERMINANTS OF RECALL

WILBUR SCHRAMM
Stanford University
AND WAYNE DANIELSON
University of Wisconsin

THIS is a study of the way in which reference groups affect how human beings handle the communications they receive. More precisely, it is a study of a *particular* reference group—the audience we expect to have to talk to concerning a given topic—and how our image of this group affects our handling of messages related to the specific topic.

Fifty-five years ago Cooley (1, especially pp. 58–62) developed the concept of "imaginary conversations." Much of our mental life, he said, is lived as a rehearsal of imaginary and anticipated conversations with persons who are important to us—or rather, with our images of those persons. "The mind," he said, "is not a hermit's cell, but a place of hospitality and intercourse. We have no higher life that is really apart from other people. It is by imagining them our personality is built up."

Many scholars have speculated as to just what effect this kind of mind life must have on perception, learning, and attitudes, but little or no experimenting had been done on it until in 1951. Grace (2) published a paper based on current developments in the theory of reference group and social perception. He showed subjects a large array of objects which ranged from intimate feminine to intimate masculine apparel. The experimental subjects were told to remember all these articles they could because they would later have to recall and repeat what they remembered. Some of them were told they would have to repeat them to a woman; others were not. *All* of them *did* report to a woman. Therefore the amount and nature of their recall could not be attributed to differential social constraint at naming intimate articles to a woman. There was no significant difference between the total number of items recalled by the two groups of subjects. However, those subjects who had been told they were to report to a woman tended to recall "feminine" articles significantly earlier in their lists. The experimenter's supposition was that these female-related items had been made more salient by the knowledge that the anticipated audience was to be female.

In 1956, Zimmerman and Bauer (3) reported an ingenious experiment which found clearer effects of an anticipated audience. Miss Zimmerman appeared before several classes, representing herself variously as speaking for an organization interested in improving teachers' salaries and as speaking for an organization concerned with saving taxpayers' money. She said her organization was seeking speakers to address its members on the subject of teachers' pay, and was also interested in how well people could remember arguments on the subject. She read a short passage to the classes, in half the classes favoring and in the other half opposing increases in teachers' salaries. She then asked the students to reproduce this passage, in writing, as accurately as possible. One week later she returned and asked them to write the passage again, as closely as they could remember it. Results were that at the end of a week, subjects remembered significantly more of the passage when they had been given a selection that was congruent with the attitude of the anticipated audience. That is, if a subject had been given a passage that opposed teachers' pay raises and was told that he was to write a speech for the Economy League, he remembered more of the passage than another subject whose passage *favored* pay raises and who was also to speak to the Economy League. The experimenters also found that subjects especially concerned with the audience (e.g., journalism students) showed more effect of this anticipated audience than subjects chiefly concerned with the topic (e.g., education students).

The present experimenters were planning a study of this problem at the same time as Miss Zimmerman was collecting her data. When they found out about her work they modified their design somewhat in order to replicate the essential feature of the Zimmerman-Bauer experiment and also to include an aspect of the relationship that was not included in the earlier work: the effect of an anticipated audience on attitude. The latter aspect of this study is not reported here because the results (which did not confirm any simple relationship of attitude change to anticipated audience) call for further research before they can be interpreted confidently.

METHOD

This experiment, then, was based on the supposition that a large percentage of the communications we receive during life are such as to call up an anticipated audience—that is, some reference group or individual to whom we think we may have to retransmit the message, or with whom we may want to talk it over, or with whom we may conduct an imaginary conversation concerning it. Given that situation, it was hypothesized that a person tends to remember the parts of a received communication which he believes will appeal to the interests or be consonant with the relevant attitudes of the anticipated audience.

In order to test this hypothesis, approximately 200 Stanford speech students were divided into three groups. A control group received attitude tests as below, but no experimental communications. The other two groups were given a two-sided set of arguments for and against lowering the voting age to 18. These arguments were equal in number for each side and were matched by judges. These test groups were told that they were to prepare speeches in competition for a prize. One test group was told that the prize was to be given by "The Society for Constitutional Government—dedicated to preserving the U. S. Constitution unchanged" and hence opposed to any change in voting age; the other that the prize was to be given by "The Association for Lowering the Voting Age to 18—If they can fight, they can vote." To reinforce this setting, the name of the mythical organization for which the student was expected to write and speak appeared on the page with the arguments, although the same set of arguments was used under both names.

TABLE 1

RECALL OF ARGUMENTS IN RELATION TO AN ANTICIPATED AUDIENCE

	Number of Arguments Recalled *For* (Mean)	Number of Arguments Recalled *Against* (Mean)	Mean Difference	Variance of Difference
Audience *For* Lowering Voting Age ($N = 76$)	3.57	3.12	0.45	1.83
Audience *Against* Lowering Voting Age ($N = 80$)	3.80	4.19	−0.39	1.64

CR for net difference (DF − DA) equals 3.98; p equals .0008 (2-tailed test).

An attitude test was administered first. Then the students were told about the prize contest, given the arguments to read, and told to remember them because they would be useful later when they would write the speech. One week later, the teacher administered a memory test on the arguments, and also repeated the attitude test. Later the students did write speeches on this subject, and small prizes were given for the best ones.

The groups were randomly assigned and found to be closely equated on the obviously relevant measures—age, year in college, sex, and major subject.

RESULTS

It was found that the subjects remembered significantly more arguments on the side they were supposed to address, as Table 1 shows. Thus it would appear that the recall of arguments from a two-sided group of arguments is significantly greater after one week for those arguments which agree with the supposed position of an audience with which the subject expects to have to discuss the topic. The hypothesis, and the Zimmerman-Bauer conclusion (3) must be regarded as confirmed.

REFERENCES

1. COOLEY, C. H. *Human nature and the social order.* New York: Scribner, 1902.
2. GRACE, H. A. Effects of different degrees of knowledge about an audience on the content of communication. *J. soc. Psychol.*, 1951, **34**, 111–124.
3. ZIMMERMAN, CLAIRE, & BAUER, R. A. The influence of an audience on what is remembered. *Publ. Opin. Quart.*, 1956, **20**, 238–248.

Received August 13, 1957. Prior publication.

ERRATA

In the article, "The Relationship Between Advanced Age and Mental Abilities", b Charles R. Strother, K. Warner Schaie, and Paul Horst, which appeared in the Septembe 1957 issue of this Journal (*J. Abnorm. Soc. Psychology*, 1957, **55**, 166–170), the cuts fo Figs. 1 and 3 were erroneously transposed in printing. The graph labelled Fig. 1 is derive from the data shown in Table 4 and represents the *corrected* T-score values. The graph labelle Fig. 3 is derived from the data in Table 1 and shows the *uncorrected* values.

In the article by E. H. Schein, W. F. Hill, H. L. Williams, and A. Lubin, "Distinguishin Characteristics of Collaborators and Resisters Among American Prisoners of War," this Journal, 1957, **55**, 197–201, the final sentence of the first main paragraph on p. 199 should read "No comparisons between resisters and *collaborators* are shown, as none of these were significant.'

THE MEASUREMENT OF PRECONSCIOUS PERCEPTION IN DREAMS AND IMAGES: AN INVESTIGATION OF THE POETZL PHENOMENON[1]

HOWARD SHEVRIN AND LESTER LUBORSKY

Menninger Foundation

IN a pioneer laboratory investigation of dream imagery, Poetzl (16) discovered that his subjects dreamt about the originally unreported parts of a previously exposed picture. He flashed colored slides of everyday scenes at 1/100″ and asked subjects to describe and draw what they saw. He then told them to record any dreams they had that night. When he asked his subjects on the following day to describe and draw their dreams, he observed that the dream imagery borrowed extensively from the unreported portions of the picture. Poetzl's subjects confirmed his observations by identifiying elements of the picture of which they were previously unaware as parts of their dream imagery.

Poetzl looked upon his findings as supporting the view that "images of indirect vision, errors of tachistoscopic vision in normal persons, visual dream-images and the visual mistakes of patients with agnosia agree with each other in many qualities regarding form and contents" (16). He attributed these similarities to a functional equivalence between brain damage to visual areas and the impoverishment of normal perception produced by tachistoscopic exposure or peripheral vision. In a brief exposure, involuntary eye movements from the point of fixation to other areas of the stimulus cannot be executed. The time limitation erects a "motoric barrier" that permits a clear perception of only part of the picture; the remainder of the picture registers in the visual apparatus but is not represented in consciousness, analogous to an exposed film that has not been developed. According to a "law of exclusion" formulated by Poetzl, only the registered but consciously unrepresented parts of a stimulus appear in the dream. The fully "developed" perception, or "print," is excluded from the dream process.

Both Poetzl and Freud were aware of the parallel between Poetzl's findings and Freud's theory of dreams. The "indifferent" perceptions that Freud identified as day-residues present in the manifest content of dreams, Poetzl equated with the unreported parts of the picture which were later dreamed about. Poetzl further assumed that there was a functional equivalence between his concept of the "motoric barrier" and the concept of a repressed wish. He maintained that the dream lifted this "motoric barrier" and permitted the successive development of the parts of the picture originally unseen, much as the dream lifted the barrier of repression and permitted the symbolic and disguised expression of unacceptable wishes. The dream work undertook its own kind of "dark room" procedure, developing the exposed but heretofore undeveloped and unprinted "film" according to its own laws rather than the laws of the waking state.

Recently, Fisher (4, 5, 6) revived general interest in Poetzl's work by providing examples of Poetzl's findings on the basis of which he suggested modifications of Freud's theory of dream formation. Klein and his associates (9, 18) felt the need for a more exact method than Poetzl's procedure, and using an ingenious "masking" technique, have attempted to show that an apparently unseen stimulus affects the perception of a simultaneously perceived stimulus. The work of these two investigators provided the immediate stimulus for the present study. Like Fisher, however, we felt the need to use the opportunities offered by the Poetzl method for tracing the vicissitudes of preconscious percepts from the time of their registration to their emergence in dreams. In a previous paper we reported on a preliminary examination of the Poetzl phenomenon together with an attempt to understand it theoretically (12).

If the basic Poetzl observation could be systematically confirmed and if a measure of delayed recall in dreams could be developed, a new approach to the experimental study of unconscious phenomena would be made possible. This method might successfully handle some of the major difficulties encountered in

---

[1] We wish to thank our colleagues for critical reading of the manuscript: Gardner Murphy, Riley Gardner, Philip Holzman, Richard Siegal, and Sidney Smith.

the many studies concerned with measuring unconscious phenomena. These investigations—carried out under such headings as subception (11), and perceptual defense and vigilance (3, 10, 15)—by and large, have utilized indirect indices such as variations in threshold time. As Klein and his associates have pointed out, in these studies ". . . non-conscious perception and the defensive reaction to it [are] both inferred from the deviant recognition times. The threshold technique of these studies . . .[provides] little if any information concerning the processes that determine the threshold measure" (9). These studies focus mainly on "defensive reactions to subliminal perception" and not upon the ways in which "subliminal registration affects conscious thought" (9).

Klein *et al.* have taken an important step toward dealing with the problem they pose. Their results suggest another question: what is the *fate* of the preconsciously registered impressions after the moment at which they affect simultaneously evoked conscious impressions? The technique discovered by Poetzl and further developed by Fisher is perhaps best designed to answer this question, for the dream would appear to bring to the surface many of these preconscious perceptions.

The Poetzl procedure has been repeated in several studies (1, 4, 5, 6, 13, 14) but thus far no measure of preconscious perception has been developed. Our primary aim in the present report is to construct a reliable measure of preconscious perception as elicited by the Poetzl technique and then to examine evidence bearing on its validity.

METHOD

Our procedure was essentially the same as the one used by Poetzl and Fisher. They varied conditions unsystematically in what were principally clinical approaches (8). We have attempted to collect our data in a systematic fashion, so that statistical procedures could be applied to the data.

We used the same picture for all *S*s. It provided a baseline from which could be assessed the nature of the "picture pull," as well as individual differences in response to the same stimulus. None of the other experimenters (1, 4, 5, 6, 13, 14) used one picture for a sizeable number of *S*s.

A single description and drawing of a complex picture may not capture all of what the *S* has consciously perceived. Our instructions therefore emphasize that the *S* should describe and draw what he saw as fully as possible. In addition, we asked for a second full description and drawing immediately following the first. At the next session (the following day) we obtained a third sample of the conscious memory of the picture. In order to make certain that nothing would be missed, a tape recording was made of the experimental sessions.

We became convinced of the importance of using a complex picture with many objects to assure that even *S*s who reported a great deal initially would have a sizeable remainder to recover in their dreams. This would make possible a statistically adequate range of recovered items.

It would be foolhardy to expect dreams to occur and be reported on request. We decided to obtain from those *S*s who did not report dreams something similar to a daydream. We asked these *S*s to describe and draw the first mental picture that came to mind. A similar technique was employed by Allers and Teler (1) and more recently by Fisher (5).

The 27 *S*s who took part in this study had participated in other experiments on perception at The Menninger Foundation. Our initial explanations of the nature of the experiment could be brief: "This is a study of perception, or how people see the world. It is similar to some of the studies you've taken part in before." The verbatim experimental instructions follow:

Part I (First Session)

1. On the count of three I'm going to flash a picture on the screen. It will appear just about where the x-mark is. [The experimenter also indicates the size of the area in which the picture will appear on the screen.] Ready: one, two, three. [The picture is flashed on the count of three.]

2. Now, describe what you saw as fully as you can. Make a drawing of what you remember seeing and continue to describe it while you draw it.

3. [The *S* is asked to describe and draw the picture again, as in #2 above. After completing the drawing and description, he is asked,] Is there anything else you remember?

4. Would you remember and write down upon awakening any dreams that you have tonight? Please bring them with you when you come tomorrow. [If the *S* asks, "What if I don't dream?" he is told, "If you keep a paper and pencil handy we've found that this helps a great deal." If he then asks, "But what if I don't dream at all?" the experimenter replies, "Even people who don't usually dream find that this helps."]

Part II (Second Session)

5. Tell me your dream (dreams). Draw and describe what was pictured in your dream, what you saw in your dream.

Alternate instructions for Ss who do not report dreams: What were the last thoughts you had before falling asleep last night? [This is followed by instruction #6 below. The remainder of the instructions, from #7 on, apply to these images as if they were dreams.]

6. Close your eyes, let a picture come to mind, let me know when you have one. . . . Fine, tell me about the picture that came to mind. Now draw what you saw and describe it as you draw it. [Two pictures are elicited in this way.]

7. [Associations are obtained to the dreams and images by means of the usual questions, "What comes to mind about the dream?" or "What comes to mind?"

—about any part of the dream which the experimenter has reason to think is likely to be important.]

8. Do you see any connection between your dream (image) and the picture you saw yesterday?

9. Now describe what you saw yesterday as fully as you can. Make a drawing of what you remember seeing and continue to describe it while you draw it.

10. I'm going to expose the picture again. [The experimenter exposes the picture on the screen and lets it stay there for the rest of the session.] What connections do you see between your recall of the picture and the picture on the screen? This time you may inspect it at your leisure.

11. What connections do you see between your dream (image) and the picture on the screen? [After this general inquiry, specific questions are asked about parts of the picture.]

12. Do you see any connections between the dream (image) and anything in the room or anything you might have seen on the way down here yesterday?[2]

13. How does this dream compare with most other dreams you've had?

14. What do you think we have been trying to get at in this experiment?

Apparatus

A slide projector with attached shutter was used. The projector was a La Belle, Model 500 with a 500-watt bulb. The shutter was an Ilex, No. 4 Acme with a range from 1 to 1/150″. Electronic calibration of the shutter was obtained for the exposure time of the experiment, 1/50″.[3] The screen used was a piece of white poster cardboard 44″ x 28″, and mounted on plyboard. The screen was hung with the long axis parallel to the floor. The room was approximately 9′5″ x 10′8″. The *S* sat against the long wall approximately 9′ from the screen. The pre-exposure field was measured at .005 foot-candles during the daytime and .000 foot-candles at night. At the moment of exposure the illumination was .200 foot-candles. With the room lights on the illumination was at .800 foot-candles. During the pre-exposure period approximately one minute was allowed for adaptation to the new level of illumination. An electronic photometer, Model 500 with a tube D was used for all measures of illumination. The size of the slide on exposure was approximately 21″ x 15″ and appeared on the screen so that the mid-point of the slide was approximately 7½″ above eye level. A black and white reproduction of our colored slide, "Inside Ancient Roman Wall,"[4] can be found in our earlier article (12). A penciled x-mark approximately ½″ x ½″ was drawn on the screen so that it would fall at the exact center point of the picture. *S* was told that the picture would appear "just about where the x-mark is" and the approximate size of the exposed picture on the screen was indicated. *S*'s head was not held stationary. On exposure of the picture the x-mark did not obscure any objects because it fell on a dark green tree in the center of the picture.

A Measure of Recall of Preconscious Perceptions

Basic to our understanding of the Poetzl phenomenon is the assumption that more of the stimulus registers than the *S* can report to us. We shall refer to those elements of the stimulus that register in the *S*'s perceptual system and are not consciously recalled, but later appear in dreams and images, as preconscious. We say recall rather than perception because in its rigorous meaning perception refers to the experience of a stimulus present to the senses; the stimulus in our experiment is only perceptible for 1/50″. It would be more accurate to refer to our *S*s' reports as the recall of a quite brief perception—a recall in which memory and judgment have already played a role. The dreams and images could also be put in this same frame of reference as another mode of recall, much as we would find in experiments on the hypermnesic character of hypnotic states (17). The recall of the stimulus present in the dreams and images differs from the conscious recall of the stimulus in that the *S* does not have the conscious intent to recall parts of the stimulus and is usually not aware that he is recalling something already exposed to his view. We shall refer to recall present in dreams and images as *unintentional recall* in contrast to the *intentional recall* obtained immediately after the exposure of the stimulus and on the *S*'s return the following day. (See Table 1.) (By referring to recall in dreams as "unintentional" we do not imply that there are no intentions other than the conscious intent induced by our instructions.)

The *unintentional recall* in dreams and images consists of at least two different types of elements: (*a*) Parts of the stimulus already reported in *intentional recall* of the stimulus. We shall refer to these as items *recalled again.* (*b*) Parts of the stimulus previously unreported in *intentional recall*. We shall refer to these items as instances of *preconscious recall.*

As a residual category, we should expect to find in the dreams and images items not present at all in the stimulus. We shall refer to these as *unrelated items.*

Once we view the Poetzl phenomenon as a function of recall we must design the measure of *preconscious recall* to take into account the effects of memory, both as a factor producing a decrement in the amount available to recall and as a factor introducing new elements in the form of additions, distortions, assimilations, etc. According to Freud's theory of dreams, we should expect these additive and distorting effects of memory to be more prominent in dreams and images than in *intentional recall*. As a way of taking into

---

[2] This instruction is one of the additions introduced by Fisher (4).

[3] We had intended using the same 1/100″ time exposure as Poetzl and Fisher, but found by electronic calibration of our shutter (Ilex, No. 4 Acme, Ilex Optical Co., Rochester, N. Y.) that the actual time exposure when set at 1/100″ was 1/50″. (By courtesy of Charles Snyder.) These types of shutters are inexact, but our time exposure proved consistently to be around 1/50″. Thus, we succeeded in our aim of providing each *S* with about the same stimulus conditions in terms of length of exposure. This possible difference in time exposure from that of the Poetzl and Fisher experiments (none of the earlier experimenters report on exact calibration) may not be crucial, as is suggested by the work of Fisher (4), and Malamud and Linder (14), both of whom obtained the phenomenon with exposures of half a minute.

[4] National Picture Slide Co., N. Y. C., Slide #9, Italy, 80.

TABLE 1
TIME SEQUENCE OF EXPERIMENTAL PROCEDURES

Experimental Procedure	Time after Exposure
Exposure of picture	0
First intentional recall	immediately following exposure
Second intentional recall	3–5 minutes
Dream (unintentional recall)	10–24 hours
Image (unintentional recall)	24–30 hours
Third intentional recall	24–30 hours

account the transformations in perceptual structure produced by memory, we decided to identify instances of *preconscious recall* in terms of a conceptual similarity between dream and picture elements. Thus, if *S* omitted the trucks from his intentional recall of the picture but dreamed about cars, wagons, or bicycles, we would credit him with the preconscious recall of the vehicles in the picture. We assume that the meaning of the percept—in this case "mode of transportation"—is more constant than any of its visual components, which might have been affected by the particular memories of the individual.[5]

The actual score for *preconscious recall* received by the *S* represents the number of elements in his dream or image having a conceptual similarity to elements of the picture which *S* did not describe or draw in his three intentional recalls. The scores for the other measures (amount of intentional recall, number of items recalled again, and number of unrelated items) were also based upon conceptual categories rather than perceptual identities.

RESULTS

Reliability of the Measure of Preconscious Recall

From the *S*s' descriptions of the picture, we made up a list of 34 major items. The major elements of the picture to which our *S*s most frequently responded were the sky, the statuary on top of the building, the "tower" (facade of the building), the building itself, the trees and shrubbery on either side of the building, the wall on which the building stood, and three vehicles on a roadway at the base of the wall. A few of the reported items were indistinct and infrequently seen and some of them were subdivisions of these principal elements. For example, the building was not only scored as a building, but windows and doors were scored separately when they were drawn and named by the *S*.

Two judges worked independently in scoring the dreams and images according to the presence or absence of the 34 items. The judge was only required to decide whether a listed item was recognizably present in the descriptions and drawings. "Judge 1" represented the pooled judgments of the two experimenters, and "Judge 2"[6] was the statistician who did most of the analyses reported in this article. We were concerned with three scores for each *S*: (*a*) the amount of *intentional recall*; (*b*) the amount of *preconscious recall*; and (*c*) the amount *recalled again*. The reliability coefficients (Pearson *r*'s) for these three variables were respectively: .92, .84, and .95.[7]

Establishing the Relationship of the Measure to Preconscious Perception

Like Poetzl, we found that parts of the picture which were not reported as seen later appeared in the dreams. Out of an average of about ten items appearing in dreams, three had not been reported previously (see Table 2).

---

[5] Another approach to the problem of identifying instances of *preconscious recall* is to conceive of a preconsciously registered percept as a dissolving photographic image whose fragments become incorporated into the imagery of dreams. From this point of view, there need not be any conceptual similarity between *preconscious recall* and the unreported aspects of the stimulus but only a fragmentary perceptual identity. Fisher relied primarily on this method of identifying *preconscious recall*. There probably is a correlation between these two aspects of *preconscious recall*. In our previous illustration concerning the omitted trucks, the conceptual similarity between the dream and the stimulus is clearly that of a mode of transportation or vehicle; but there is also a perceptual identity present insofar as an *S* usually draws his wagons with wheels, fenders, etc. Our decision in favor of a measure based on conceptual similarity was a first step dictated by the need to achieve adequate reliability, which seemed more readily obtainable by judging conceptual similarities than by judging minute, subtle, and fragmentary perceptual identities.

[6] Lolafaye Coyne.

[7] The score given each *S* by the two judges could be based on different items. We therefore determined the amount of agreement between Judge 1 and Judge 2 for all *items*. This is the same as asking how well the judges agreed on whether an *S* was drawing and describing, for example, a building, a car, or a tree. The percentage agreement between the two judges for all items, pooling all *S*s, was 68. Almost all the disagreements had to do with one judge scoring an item and another not scoring it. There were two main reasons for these disagreements. Occasionally, one judge noticed an ambiguous item and decided it did not represent the object, while the other judge decided that it did resemble a recognizable object. A more prevalent type of disagreement was essentially one of oversight by one or the other of the judges of an unambiguously represented item. The descriptions were rated more unreliably than the drawings. When oversights were called to the attention of the judges they usually immediately confirmed having overlooked the object.

We also found that another source of *unintentional recall*, images, was effective in eliciting new picture items. Indeed, the images proved to be significantly more productive of *preconscious recall* than the dreams (see Table 2), a finding that raises some interesting questions beyond the scope of this paper. We have labeled these new items in the dreams (and images) *preconscious recall*. How can we show that this is a proper description?

If we could put "tracers" on the picture items analogous to radioactive isotopes, we could follow them through the mental apparatus and our task would be relatively simple. Until such psychologically radioactive isotopes are discovered, however, we must learn about the validity of the measure by bringing together evidence that it behaves in accordance with our understanding of preconscious perception. We explored hypotheses based on the assumption that preconscious perception occurred and was utilized in dreams and images.

Poetzl's "law of exclusion." Poetzl stated: ". . . out of the total effect of a visual situation the once-performed part [i.e., the part which appeared in consciousness] disappears in the after-effect, but those preconsciously grasped parts, which were not yet developed, continue to be effective. If the grasp of consciousness is at its maximum, the latent after-effects in preconsciousness are at a minimum and vice-versa" (16). This "law" might lead one to expect that the more items the *S* reported (*intentional recall*), the smaller the number of items left to be "developed" in the dreams and the images (*preconscious recall*); similarly, the greater the *S*'s "conscious grasp" of the picture (*intentional recall*), the less the number of items *recalled again* in the dreams and images. Our findings indicated otherwise. The correlation between *intentional recall* and *preconscious recall* was not significant (−.06), while the correlation between *intentional recall* and *recalled again* was significant, but in the opposite direction from that predicted (.69, $p < .001$). Poetzl therefore was wrong if the two predictions followed from his "law of exclusion"—the amount one sees initially of a picture is unrelated to the amount one recovers of it in dreams and images; and the more one sees initially, the *more* often the same things will be seen again in the dreams and images.[8]

TABLE 2

MEAN NUMBER OF ITEMS RECALLED IN EACH CATEGORY FOR DREAMERS AND NONDREAMERS

Category of Recall	Dreamers (N = 16)	Nondreamers (N = 11)	t
Preconscious recall	2.9	4.7	2.25*
Intentional recall	9.3	9.5	0.17
Items recalled again	2.4	3.6	1.09
Unrelated items in unintentional recall	4.7	5.5	0.44

* $p \leq .05$.

In view of these negative results, one may wonder whether we are measuring the same phenomenon that Poetzl observed. There is, however, another way of testing Poetzl's "law of exclusion" made possible by our obtaining three intentional recalls of the picture, rather than one, as in previous studies. Quite frequently we noted that items not in the first recall emerged in the second or third recalls. We were led to ask whether more new items would appear in the dreams (and images) than in successive intentional recalls. Possibly Poetzl's "law" was intended mainly to emphasize that the dream is a better vehicle than *intentional recall* for *preconscious recall*, rather than to imply that the dream is the exclusive vehicle for preconscious perception. Our definition of *preconscious recall*—the appearance in dreams and image of picture items not previously reported in *intentional recall*—might therefore have been too limiting. Perhaps the new items in successive *intentional recalls* could be viewed as the feeble workings of *preconscious recall* under the adverse conditions of the waking state. Similarly, perhaps the dreams and images are relatively inefficient means as compared to *intentional recall* for recalling conscious perceptions. We, therefore, restated our two predictions:

---

[8] The relationship between *intentional recall* and *preconscious recall* could be obscured by an uncontrolled factor—we have no way of knowing how much of the picture actually registered in our *S*s' visual apparatus. There may be wide individual variations in how much a person can "receive," adding together conscious and preconscious perception. Visual acuity might be one such variable which would place a physiological limit upon how much could be perceived at any one time. An analysis which pitted the *S* against himself would control for inter-individual variations in visual acuity and other limiting physiological conditions.

The relationship between *intentional recall* and the number of items *recalled again* in dreams and images is also limited by an uncontrolled factor: on the basis of chance expectations the more things a subject saw in the picture consciously, the greater number of things he could recall again in his dreams or images. If a person reported seeing ten items, he could see again in his dreams a maximum of ten items; but a person who saw only two items could only recall a maximum of two in his dreams. In this sense, a positive correlation was "built into" our measure. If the *S* were compared with himself under the conditions of *intentional* and *unintentional recall*, the amount consciously reported would be eliminated as a source of individual variation.

1. Each *S* should recall significantly more new items (*preconscious recall*) in his dreams or images than in his second and third efforts at *intentional recall*.

2. Each *S* should recall significantly fewer old items (*items recalled again*) in his dreams or images than in his second and third efforts at *intentional recall*.

Both predictions were borne out by our results. (See Tables 3 and 4.) For our analyses, a difference score was obtained for each individual between his performances in *intentional* and *unintentional recall*, using the first intentional recall as a baseline. In analyzing *preconscious recall*, for example, the difference was obtained between the number of new items in *unintentional recall* (dreams or images) and the number of new items present in *intentional recall* (second or third recall, or their average, depending on the particular comparison involved). These difference scores were averaged for all *S*s and the significance of the average difference tested. *Intentional recall* was a better vehicle for retaining conscious perceptions, but could not measure up to the capacity of *unintentional recall* to activate preconscious perceptions. These relationships applied to dreams and images taken separately or combined. The findings support our restatement of Poetzl's assertion that dreams are prone to "exclude" conscious perceptions in favor of preconscious perceptions. And, our results indicate that this "law" can be generalized to apply to a type of "forced fantasy," the images.

TABLE 3

COMPARISON OF PICTURE ITEMS PRECONSCIOUSLY RECALLED IN UNINTENTIONAL AND IN INTENTIONAL RECALL

Comparison	N	Mean Difference	σ_D	t
Unintentional recall vs. second intentional recall	27	2.3[a]	2.75	4.35****
Unintentional recall vs. third intentional recall	26	2.5	2.40	5.50****
Unintentional recall vs. second and third intentional recall combined	26	2.4	2.52	4.93****
Second intentional recall vs. third intentional recall	26	0.3	1.55	1.03

**** Significant at $p < .0001$.

[a] A positive difference is in favor of the first variable mentioned in the comparison.

Freud's hypothesis about "indifferent" perceptions. We wondered what characteristics of preconscious perceptions made it possible for them to play a role in dreams and images. In his theory of dreams, Freud identified preconscious perceptions as those which are ". . . indifferent and thus had no attention paid to them; . . . or have been rejected and have thus had attention promptly withdrawn from them" (7). In either case, these preconscious impressions "have not yet been requisitioned by waking thought-activity; and for reasons of censorship . . . [the dream process] transfers psychical intensity from what is important but objectionable to what is indifferent" (7). In this last statement Freud suggested a reason for the operation of Poetzl's "law of exclusion." Fully developed conscious perceptions cannot so easily be used as a "cover" for unacceptable thoughts.

We would predict then that preconscious perceptions, as they emerge in dreams, would

TABLE 4

COMPARISON OF PICTURE ITEMS RECALLED AGAIN IN UNINTENTIONAL AND INTENTIONAL RECALL

Comparison	N	Mean Difference	σ_D	t
Unintentional recall vs. second intentional recall	27	−2.9[a]	2.49	−6.02****
Unintentional recall vs. third intentional recall	26	−3.0	3.03	−5.16****
Unintentional recall vs. second and third intentional recall combined	26	−3.0	2.67	−5.81****
Second intentional recall vs. third intentional recall	26	−0.0	1.90	−0.11

**** Significant at $p < .0001$.

[a] A negative difference is in favor of the second variable mentioned in the comparison.

be associated with a strong affective charge displaced from ideas originally intensely cathected to ideas having only a weak charge. Moreover, because these intensely cathected ideas are the derivatives of unacceptable unconscious wishes, we should expect the affective charge to be unpleasant in nature. With these considerations in mind, we predicted that *preconscious recall* would be associated with unpleasant affect.

In order to check this prediction, we devised a measure of unpleasantness using a five-point scale in which two judges rated the maximum point of unpleasantness attained (*peak unpleasantness*) within the dreams, images, and the three reports of *intentional recall*. One judge was not aware of the expected relationship between *peak unpleasantness* and *preconscious recall*. We did not get the *S*s' own rating of the dream's unpleasantness. Illustrating a low *peak unpleasantness* rating would be a dream in which there was no more unpleasant part than a man reading a book; illustrating a high rating would be a dream in which a woman was sitting in her car while terribly frightened by a hail storm.

Confirming our prediction that *preconscious recall* would be associated with unpleasant affect, we found a positive correlation between *peak unpleasantness* and the amount preconsciously recalled. Both for our group of *S*s as a whole (.57, $p < .01$) and for the dreamers (.75, $p < .001$) and nondreamers (.60, $p < .05$) the relationship was significant.

It could be argued, however, that the unpleasantness in dreams and images may not be a function of the preconsciously recalled items at all, but a function of other aspects of the dream. With this objection in mind, we correlated *peak unpleasantness* with the number of items that were *recalled again* and the number of *unrelated items* in the dreams and images. The two correlations were, respectively, .01 and .29. These nonsignificant correlations support the assumption that it was the preconsciously recalled items and not the other contents of the dreams and images that were associated with *peak unpleasantness*.

An hypothesis concerning a formal attribute of the picture. Our measure of *preconscious recall* was based on the picture's content. The *formal* attributes of the picture may also be capable of *preconscious recall*. One formal property is the angularity of line produced by the shapes of the picture items.

Each drawing made by an *S* was judged according to its *angularity* (as opposed to roundedness) on a five-point rating scale. For the drawings of the picture a reliability coefficient of .88 was obtained for two judges; the reliability coefficient for the dream and image drawings was .81.

In constructing our measure of *angularity*, we assumed that the average angularity for the group of *S*s would be a good estimate of the angularity of the picture. Statistical analysis indicated that there were no significant differences among the means for *intentional* and *unintentional recall*, a finding in line with the assumption that the mean remained a relatively constant estimate of the angularity of the picture. We can consider individual deviations from the central tendency as errors in reproducing the angularity of the picture. A "z" score was obtained for each *S* representing the difference between the angularity of his *unintentional recall* drawings and the mean angularity of the total group for *unintentional recall*. We would expect that the smaller the "z" score (i.e., the closeness with which the *S* approximates the average angularity of the drawings), the greater the amount of *preconscious recall* of picture contents. The correlation between the "z" scores and *preconscious recall* of picture contents was $-.49$ ($p < .01$). This relationship was stronger for the dreamers ($-.57$, $p < .05$) than the nondreamers ($-.22$).

Preconscious recall of the picture content was shown to be significantly correlated with *peak unpleasantness*; angularity should similarly be correlated. We found a significant correlation in the expected direction ($-.40$, $p < .05$).

As with *peak unpleasantness*, however, it could be argued that *angularity* was equally related to other aspects of recall, such as items *recalled again* and *unrelated items*. Our expectation was that there would be no systematic relationship between *angularity* and these two variables. The correlation between angularity of dream and image drawings and the number of items *recalled again* was $-.09$. The correlation between *angularity* and the number of *unrelated items* present in the dreams and images was, however, significant, $-.41$ ($p < .05$), indicating that the nonpicture items appearing in *unintentional recall* are in some way related to the accuracy with which the angularity of the picture is represented in the dream and image drawings.

In view of these findings, the angularity of the picture can perhaps be treated as an aspect of *preconscious recall*. The degree to which this angularity is rendered accurately is related to *peak unpleasantness* and items *recalled again* in the same way as the *preconscious recall* of picture contents. Our measure of *preconscious recall* relied on the conceptual similarities found in the picture content. *Angularity*, on the other hand, would extend our measure of *preconscious recall* to perceptual identities between the dream and picture.

Evidence Against the Chance Occurrence of the Picture Items in Dreams and Images

One disadvantage of our measure of *preconscious recall* derives from the range of items that can be included in a conceptual category. How certain can we be that the wagons seen in the dream, for example, are not derived from a source other than the trucks in our picture? And when we note that most of the objects in our picture are commonplace, the strength of this criticism is increased. If we could, however, select from our pictures the items that occur rarely in everyday experience and demonstrate that these rare items are preconsciously recalled more frequently than would be expected by a standard of chance occurrence, we would have presumptive evidence that the source of *preconscious recall* was our picture and not an extraneous stimulus. Comparison data from the literature on extrasensory perception may help establish a standard by which to judge frequency of occurrence of percepts. Catalogues have been constructed to show how frequently an *S* will guess an unseen object, selected at random and without his knowledge. A catalogue has been made for British *S*s (2), and now an unpublished version for American *S*s by Taves (19) (available by courtesy of Gardner Murphy) of 8,164 items produced by 272 *S*s.

The Taves catalogue can help us answer this question: are the items in our picture that have a low frequency of occurrence in the Taves catalogue and are presumably uncommon in everyday experience recalled preconsciously more frequently than their rare occurrence in the Taves catalogue would lead us to expect? If we find, for example, that an item in our picture which occurs only two per cent of the time in the Taves catalogue, but occurs 20 per cent of the time in our *preconscious recall*, it would seem likely that the picture and not some extraneous source contributed that item. When we selected those items in our picture which made up two per cent or less of the total number of responses in the Taves catalogue, we found that when added together those same items occurred much more frequently in preconscious recall. The chi square was at the .008 level of significance. (If we selected items which occurred .05 per cent or less of the time in the Taves catalogue, the results were even more significant—the chi square was below the .001 level of significance.)

The comparison with the Taves catalogue supported our assumption that nonexperimental sources alone cannot account for *preconscious recall*. For future work, however, it would be helpful to compile item frequencies for dreams and images, so that the standard of chance occurrence might be more closely related to our data than the ESP catalogues.

DISCUSSION

Our findings form a pattern consistent with the assumption that preconscious perception exists and can be elicited and measured. We have developed a measure, *preconscious recall* in dreams and images, which behaves in accordance with an understanding of preconscious perception. Our *S*s' dreams and images contain significantly more originally unreported items than do successive efforts at *intentional recall*, while retaining fewer of the originally seen parts of the stimulus. An estimate of unpleasant affect is associated with the new items appearing in the dreams and images. These two findings support Poetzl's "law of exclusion" and Freud's conception of preconscious perception as the raw material "preferred" by the dream. The finding concerning the *angularity* of the drawings indicates that the measure of *preconscious recall* can perhaps be extended to such formal properties of the stimulus as shape and position. A method was successfully applied for assessing the "chance" occurrence of picture items in dreams and images. The rare items in the picture appear significantly more frequently in dreams and images than a standard of expected chance occurrence (i.e., Taves catalogue). Taken together these four findings strengthen the conviction that Poetzl's technique for relating dreams to specific and identifiable daytime stimuli is an effective method for the experimental study of dreams.

Much more remains to be understood about the phenomenon. For example, why are the *unrelated items* unassociated with unpleasant affect? Many of these unrelated items could be day residues stemming from nonpicture sources, and as such should have unpleasant affects associated with them.

It would be desirable if a method could be devised for identifying at what point in the cognitive process the percept acquires anxiety arousing connotations and becomes available as a day residue. Customarily, it has been assumed that anxiety arousing connotations are present at the moment of perception and indeed produce the perceptual omissions or distortions. The process may be the reverse—an initial ambiguity or vagueness in the percept may make it possible for anxiety arousing meaning to adhere to it at some later point.

Lastly, the focus of our measure, preconscious perception, is upon items omitted from conscious report. But many items that were reported as seen appeared again in the dreams and images. Do items that are reported as seen appear in the dream via a different process from omitted items? A thorough study of the *S*'s initial reports may reveal many different types of items which are as likely to appear in dreams and images as initially omitted items. We have preliminary evidence that doubt surrounding items is one of the most significant indications that they will appear in dreams and images.

Among other factors to be taken into account in future experimentation, it would be naive to neglect the fact that what is reproduced in *preconscious recall* is stimulated by much more than the experimenter's picture. Frequently the contents of the experimental room and the building become part of the "experimental stimulation." There was at least one striking example of this. The masonry which was occasionally correctly located in the wall below the building in the picture could as easily refer to the walls of the office in which the experiment took place, which were of unplastered concrete blocks.

It could be argued that the *amount* of *preconscious recall* is a function of the number of items the person reports in his dreams and images. If the *S* talks long enough he may mention many of the common items included in our picture. The simplest way to test for this is to count the number of items unrelated to the picture appearing in the dreams and images and to correlate this sum with the sum for *preconscious recall*. The correlation for the entire group was not significant (.29).

Several methods will be profitable for future studies. (*a*) *Prediction study:* By predicting from the *intentional recall* of the picture what items will appear in a dream (or image), we have another way of exploring the phenomenon. By obtaining more than one complete description and drawing of the picture in quick succession, predictions are facilitated. We observed that changes from the first to second set of drawings and descriptions were often predictive of what would appear in the dreams and images. Predictions have been tried on an unsystematic basis with encouraging success (12). (*b*) *"Own Control" method:* Images can be elicited both before and after a stimulus picture is exposed. A comparison of the images produced before seeing the picture with those produced afterward can point to the influence of the picture as compared with the readiness for imagining nonpicture items. This procedure has also been tried unsystematically (6). (*c*) *Experimental pictures containing unusual objects:* As the Taves frequencies indicated, most of the objects in our picture are common. It is important to use a picture in which many or all of the objects are rare. The frequency with which the objects are commonly associated with each other should also be controlled. Instances of *preconscious recall* might really be commonplace associations—car and street, shrubbery and house, building and windows, etc. The new items in the dreams and images could be *associations* to the originally seen items, dream continuations of thoughts begun in a conscious state. (*d*) *Intra-individual consistency in response to a number of pictures:* By using a number of different pictures for the same person we can separate effects due to individual differences from the effects produced by picture content. The other three methods can be employed with a number of pictures as easily as with one.

SUMMARY

In a pioneer investigation, Poetzl discovered that *S*s dreamed about the originally unreported parts of a picture exposed tachistoscopically. If this discovery can be confirmed and a measure developed of preconscious perceptions as they appear in dreams, a new approach becomes available for the experimental study of unconscious phenomena.

A kodachrome slide was exposed at $\frac{1}{50}''$ to 27 *S*s. They were asked to describe and draw the picture three times and then to record any dreams they had that night. From *S*s who did not report dreams an "image" was obtained by asking them to close their eyes and describe and draw the first picture that came to mind.

High interjudge reliability (.84) was achieved for a measure of preconscious perception (*preconscious recall*) based on the number of new picture items appearing in the dreams and images. The meaning of the measure was explored by testing two hypotheses:

1. Poetzl's law of exclusion: Dream imagery "excludes" conscious perception in favor of preconscious perception.

2. Freud's hypothesis about "indifferent" perceptions: Freud asserted that the neutral or "indifferent" character of preconscious perception permitted it to serve as a "cover" for unconscious ideas which would not otherwise escape dream censorship. If preconscious perceptions were the ones that most often linked up with derivatives of threatening ideas, unpleasant affect would more likely be associated with the recall of preconscious perceptions than with other aspects of the dream.

Our data very significantly supported the Poetzl hypothesis for picture content and also for a formal attribute of the picture, its angularity. Consistent with our inference from Freud's hypothesis, we found a significant correlation of .57 between *preconscious recall* and *peak unpleasantness*.

These findings provide evidence for the Poetzl phenomenon consistent with the assumption that preconscious perception can be elicited and measured both in images and in dreams.

REFERENCES

1. ALLERS, R., & TELER, I. "Ueber die Verwertung unbemerkter Eindruecke bei Associatonen." *Ztschr. f. Neurol. und Psychiat.*, 1924, **89**, 492–513.
2. CARINGTON, W. Experiments on the paranormal cognition of drawings, III. *Proc. Amer. Soc. for psych. Res.*, 1944, **24**, 1–107.
3. ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–564.
4. FISHER, C. Dreams and perception: The role of preconscious and primary modes of perception in dream formation. *J. Amer. psychoanal. Assn.*, 1954, **2**, 389–445.
5. FISHER, C. Dreams, images, and perception: A study of unconscious-preconscious relationships. *J. Amer. psychoanal. Assn.*, 1956, **4**, 5–48.
6. FISHER, C. A study of the preliminary stages of the construction of dreams and images. *J. Amer. psychoanal. Assn.*, 1957, **5**, 5–60.
7. FREUD, S. *The interpretation of dreams.* (Trans. by James Strachey) New York: Basic Books, 1955.
8. KLEIN, G. Discussion remarks on Charles Fisher's "A study of the preliminary stages of the construction of dreams and images." Unpublished manuscript read at the New York Psychoanalytic Society, 1956.
9. KLEIN, G., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, S. Preconscious influences upon conscious cognitive behavior. Paper read at Annual Meeting, American Psychological Association, San Francisco, 1955.
10. LACY, O. W., LEWINGER, N., & ADAMSON, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–174.
11. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **48**, 113–122.
12. LUBORSKY, L., & SHEVRIN, H. Dreams and day-residues: A study of the Poetzl observation. *Bull. Menninger Clin.*, 1956, **20**, 135–148.
13. MALAMUD, W. Dream analysis: Its applications in therapy and research in mental diseases. *Arch. Neurol. & Psychiat.*, 1934, **31**, 356–372.
14. MALAMUD, W., & LINDER, F. E. Dreams and their relationship to recent impressions. *Arch. Neurol. & Psychiat.*, 1931, **25**, 1081–1099.
15. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
16. POETZL, O. Experimentell erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Ztschr. f. d. ges. Neurol. und Psychiat.*, 1917, **37**, 278–349.
17. ROSENTHAL, B. G. Hypnotic recall of material learned under anxiety and non-anxiety producing conditions. *J. exp. Psychol.*, 1944, **34**, 369–389.
18. SMITH, G. J. W., & HENRIKSSON, M. The effect on an established percept of a perceptual process beyond awareness. *Acta Psychol.*, 1955, **11**, 346–355.
19. TAVES, E. The construction of an American catalogue. *J. Amer. Soc. psych. Res.*, 1945, **39**, 151–156.

Received June 3, 1957.

AFFECTIVE RESPONSE TO PSYCHOANALYTIC DEATH SYMBOLS

W. W. MEISSNER
Saint Louis University

IN RECENT years, an attempt has been made to account for the anxiety reaction typical of neurosis in terms of a "fear of death" (5, 14). Psychiatrists and psychoanalysts have begun to recognize that at least some kinds of neurotic behavior can be described and treated in relation to this dread of death. In general, this fear seems to take the form in which the person dreads the dissolution of his own personality. The present study is an attempt to investigate the possible relation between the death concept and emotional response.

The measure of affectivity employed in this study was the GSR. The general evidence on the GSR as a measure of emotional response has been summarized by Woodworth and Schlosberg (19). Particularly relevant for this investigation is the evidence which suggests that the GSR is applicable as a measure of unconscious emotionality. The work of McCleary and Lazarus (8, 9) on autonomic discrimination without awareness, McGinnies' findings on perceptual defense (10, 11), and Jones's experiments on social behavior in relation to GSR responses (6) seem to indicate that the GSR is an appropriate index of internal, inhibited, unconscious emotionality. The hypothesis of this study is that the presentation of stimulus words, which have been found by psychoanalytic methods to be symbolic of the death concept, elicits unconscious emotional responses, which are indicated by a greater GSR amplitude to such words than to neutral terms.

METHOD

Subjects

The *S*s were 40 Roman Catholic seminarians with considerable training in humanities and religion. Ages ranged from 23 to 45 years, with a median of 28 years. They were all psychologically naive and had no information about the purpose or specific content of the experiment.

Apparatus and Materials

The GSR was recorded on a 110-volt, 60-cycle A.C. Garceau recording galvanometer, made by the Electro-Medical Laboratory. Circular palmar electrodes of amalgamated zinc, applied with Brunswick electrode paste, were used.

Twenty death symbol words were selected from the literature and randomly mixed with 30 randomly selected words from the Kent-Rosanoff Word List (13) to compose the experimental word association list. Ten additional words were drawn from the Kent-Rosanoff list to provide a warm-up period for the *S* to adapt to the experimental procedure. The experimental design is based on the comparison of responses to the experimental grouping of 20 death words with the responses to the 30 control words.

The same 20 death symbol words were also used to compose a written test,[1] in which the words were simply presented in the following order:

1. bird
2. journey
3. candle burning out
4. to burn
5. across the bridge
6. sleeping person
7. train trip
8. old man
9. statue
10. to drive away
11. black
12. water
13. the silent one
14. to depart
15. thunder
16. stranger
17. terminal
18. the end
19. four
20. the thirteenth

*S*s were given the written test with the following instructions, which were printed on the test blank: "This is a test of interassociation. Try to group as many of the words as possible under a common association or concept or common idea to which they refer. Then indicate the common idea, together with the numbers of the words from the list which you associate with it, in the spaces provided. Do this for each common association. The more common associations you can find and the more words you can relate to each association, the better it will be." This test was used to verify the content of the responses to these death symbol words when they were presented in the experiment.

Procedure

The experimental procedure (16) was the same for each of the 40 *S*s. The *S* was brought to the experimental room and asked to lie down on a bed provided for the purpose. He was told only to relax and rest. The rest period lasted for 15 minutes; during this period, the galvanometer was permitted to warm up. The electrodes were then attached. A small amount of nonpolarizing electrode paste was applied to the *S*'s palm, and a small amount of paste was evenly spread over the surface of the electrode. The electrode was held in place by an elastic band.

The *S*s were instructed to (*a*) respond to the stimulus word with the first word that they happened to associate with it, (*b*) reflect quietly on the content of their response, (*c*) remain as relaxed and inert physically as possible, and (*d*) reflect only on the content of their response, excluding all other extraneous ideational material. Then the ten warm-up words were presented, followed immediately by the 50 experimental words. The words were pronounced in an even and neutral tone of voice. A reading was taken on the galvanometer at the point of the presentation of the stimulus word and at the point of maximal deflection for each response. The difference between these two readings was taken as the change in resistance for that particular response. At each stimulus, *E* waited until the level of resistance had returned to a relatively steady condition before presenting the next stimulus. The interval between stimuli varied from word to word and from *S* to *S*, but the average interval was approximately 20 seconds (2). The entire procedure lasted about 50 minutes for each *S*.

---

[1] This written test was empirically constructed on another group of 20 seminarians comparable to the experimental subjects.

In the computation of the results, some of the changes from the first to the second reading which were positive instead of negative, that is, which indicated a rise in resistance instead of a drop, were eliminated by equalling them to zero. It seemed that these positive responses were an artifact of the constant positive drift observed in all *S*s. Evidently, these responses were composed of a small negative decrement which was immediately overcome by the positive drift and thus could not be recorded by *E*. This happened in only a few instances.

As soon as the list of stimulus words had been presented and the GSR responses recorded, the electrodes were removed. The *S* was then asked to write down all the stimulus words (those pronounced by *E*) that he could recall. No time limit was imposed. It was thought that a significantly higher recall of death stimulus words would indicate greater affective involvement. The differences, however, were not significant. This would seem to support Merrill's conclusions in regard to the inadequacy of recall for indicating affective involvement in word association procedures (12).

Immediately after this, the *S*s were given the written test described above. *S*s were told that they could take as much time as they pleased in filling out the test. The only other instructions were those printed on the test blank. The written test was constructed to show that the responses to the death symbol words, which were measured by the GSR, were made on the basis of death association and not on some other vicarious association. The *S*s responded by writing in the provided space a word which any or all of the death symbol words may have suggested to them. The numbers of the death symbol words which suggested this common referent were recorded together with the referent. A typical response would be: "Death—2, 3, 6, 8, 11, 13, 14, 18." There was no limit to the number of referents an *S* could give. Each such referent was scored as a response, and was given a value score equal to the total number of words associated with it. Thus, the typical response cited above would have a score of 8. The highest possible score for any one response was 20, since there were only 20 words in the list. The total group score for each such referent was obtained by adding the individual scores from each *S*. Only explicit reference to the word "death" was taken as evidence of association to the death concept.

Conditions

Conditions for the experiment were structured with Venables' findings on the effect of temperature and humidity on the GSR in mind (17). The experiment was conducted in a soundproofed, airconditioned room of constant temperature and humidity. The temperature was kept at approximately 72 degrees Fahrenheit, the humidity at 70 to 75 per cent. There were no distracting noises and no interruptions to disturb the *S*s.

RESULTS

The written test of interassociation was intended to yield an index of the tendency of these *S*s to associate the death symbol words to the death concept. In this way it was hoped that some indication could be obtained of the extent to which the GSR responses to the death symbol words were actually determined by association to the death concept. The total score of all the responses given by the 40 *S*s was 2430. Out of this total, the death referent had a total score of 267, which was higher than the score of any other referent. No direct statistical evaluation was made of the above chance expectation of the death referent score. However, a random selection of an equal number of words from the Kent-Rosanoff list provides an association with the death concept in only three out of 20,000 responses (13). A statistical comparison of these proportions is not possible due to the differences in the method of obtaining responses, but the comparison is nonetheless suggestive, if not indicative.

It was also felt that the frequency of the death association as indicated by the death score should be compared to the frequency of association to the other referents given by the *S*s. To obtain this measure, the proportion of death responses was compared by a *t* test with the proportion of that referent which had the second highest score. The referent "travel" was second highest with a score of 200, and comparison with the death score yielded a *t* of 3.265. This difference is significant at well below the .01 level and indicates that other factors than chance were operative. Thus, the death score here suggests that the *S*s to a significant degree associated the death symbol words with the death concept and that these words elicit associations to the death concept with greater probability than to any other referent.

The experimental data obtained from the GSR measurements were treated in a simple analysis of variance for matched groups or repeated observations (3, 7). Because of the high frequency of associations to the travel concept, it was felt that the analysis should be conducted on three groupings. By an analysis of the data from the written test, the 20 death symbol words were divided into those which seemed related to the travel concept and those which did not. These two groups, together with the neutral words from the stimulus list, constituted the three categories: Death-non-travel, Death-travel, and Non-Death. This design was intended to determine the significance of the differences between the amplitude of the GSR responses to the death symbol words, on the one hand, and the amplitude of the responses to non-death or neutral words, on the other, and whether the words related to the travel concept actually constituted a separate response category from the death association in the GSR measurement. A significant difference between the Death-non-travel and the Death-travel groups would suggest that the responses to each group of words was governed by a different set of associations. The results of the analysis of variance are presented in Table 1.

The significance of the differences revealed by the analysis of variance made the application of the *t* test necessary to locate the differences. The *t* was applied with the error term or residual of the analysis as the variance of the means. Results are presented in Table 2. The difference between Death-non-travel and Non-Death words, and the difference between Death-travel and Non-Death words are significant beyond the .01 level. The difference between the Death-non-travel and the Death-travel words is nonsignificant. Within the limits of this design, one may conclude, then, that the greater amplitude of the GSR response to death words than to neutral words was probably due to association to the death concept, in so far as the *S*s tend to associate these words to "death" more than to any other concept.

TABLE 1
ANALYSIS OF VARIANCE FOR GSR RESPONSES[a] TO DEATH-NON-TRAVEL, DEATH-TRAVEL, AND NON-DEATH STIMULUS CATEGORIES

Source of variation	*df*	Mean square	*F*
Between response categories	2	0.93937	24.323*
Within response categories	117	0.47797	
Between *S*s	39	1.35668	35.129*
Residual	78	0.03862	
Total	119		

[a] Calculations based on log conductance units.
* Values are significant at less than the .01 level of confidence.

TABLE 2
t TESTS APPLIED TO THE ANALYSIS OF VARIANCE FOR GSR RESPONSES[a]

Categories	*t*	*P*
Death-non-travel vs. Non-Death	6.441	<.01
Death-travel vs. Non-Death	5.538	<.01
Death-non-travel vs. Death-travel	0.903	n.s.

[a] Calculations based on log conductance units.

DISCUSSION

The results of the analysis of variance and the *t* tests applied to it indicate that the differences between the amplitude of the GSR in response to death symbol words and the amplitude in response to the neutral control words are significant. This result supports the main hypothesis of this study that death symbols elicit a differential emotional response. The results shown in Table 2 also indicate that responses to the Death-travel words do not differ significantly from those to the Death-non-travel words. There is therefore no evidence that the travel words evoked a distinguishable category of GSR response. The responses to both were apparently based on a common associative content involving the death concept. The reason for the prominence of the travel responses on the written interassociation test seems to be simply the direct implication of travel in a number of the terms.

The hypothesis behind the written interassociation test was that the *S*'s reflection on his own responses to the death stimulus words would tend to tap the complex of associative material which related to the death concept and involved an affective reaction. The presentation of the same stimulus words together in a group on the written test should elicit the same associations as did these same words in the experimental presentation. This multiple probing of the same association area on the written test should yield, then, explicit reference to the death concept if these symbolic words had any tendency to be referred to the same psychic content. The results of the written test indicate that such a tendency was manifested for these *S*s over these symbolic words. This evidence does not directly imply that the responses measured by the

GSR were actually responses involving reference to the death concept. But the tendency of these *S*s to associate these words to the death concept increases the probability that this was in fact the case.

During the course of the experiment, the infrequency of the overt response of "death" to the death symbol stimulus words was noteworthy. When death symbol words were presented, the *S*s responded with the word "death" only three or four times out of 800 opportunities. This low frequency, compared to the much higher one elicited by the written test, seems to suggest that the reaction to the death words which occasioned the greater GSR deflections may well have been on an unconscious level. This possibility fits in well with previous findings on the GSR as a measure of unconscious affectivity. The relation is at any rate highly suggestive.

The results and conclusions presented here presume that the *S*s were able to observe all the experimental conditions. No check can be made with any accuracy on the extent of interference from extraneous ideational material during the course of the experiment. *S*s were directed to exclude such material, but this is no guarantee. A claim could be made for the unconscious character of the connection between symbol and associative pattern, so that one could hypothesize that the *S* was reacting on the basis of death association. However, this hypothesis was not open to validation in this experimental design.

A factor that may have biased the results of this study is the common religious attitudes and training of the *S*s. One might expect a greater involvement in the death concept in persons for whom it has greater meaning. The extent of this influence can only be estimated from parallel studies with an equivalent design on different *S*s. Another uncontrolled factor was the tone of voice in which the stimulus words were presented. An attempt was made to pronounce them in a neutral tone, but the possibility remains that the *S*s picked up some subtle clue which may have influenced their responses.

It should be pointed out that the positive drift of the resistance level, which is commonly observed in *S*s in a relaxed condition (1), may have tended to decrease differentially the amplitude of the responses. This would have a slightly greater effect on the responses of greater amplitude, tending to reduce the greater resistance drops toward the average of the lesser resistance drops. Over the whole spread of measures, this would reduce the mean difference between the death and non-death responses.

McGinnies' demonstration of higher GSR responses to emotionally toned words (10, 11) was criticized by Howes and Solomon (4) on the ground that the emotionally toned words were less frequent in common usage and that this may have accounted for the GSR differences. This same criticism might be applied to the present design. However, checking the frequencies of the death words used in this study against the frequencies of the non-death words on Thorndike's frequency list (15) indicates that there is no significant difference in the relative frequencies.

Methodologically, the device used here to identify associative content is difficult, demanding, and in the final analysis yields nothing but indirect evidence for a plausible argument. It suggests only a tendency in the *S*s and does not identify the precise nature of their responses. Improved instrumentation may well be worth the effort of devising it. Additional light may be thrown on the connection between anxiety and the death concept by the introduction of some clinical criterion or index of anxiety. In this connection, GSR results from normal and neurotic groups might be compared. The effects of induced anxiety by the use of lysergic acid, mescaline, and other anxiety producing drugs seem to offer possibilities (18).

SUMMARY

The design of this study was directed to testing the hypothesis that there is a relation between the concept of death and affect-producing mechanisms as demonstrated by the presentation of symbolic material which is suggestive or representative of the death concept as stimuli eliciting a GSR response.

Twenty death symbol words were selected from the literature and mixed at random with 30 noncritical words to form a word association list of 50 words. These words were presented to 40 subjects in a word association procedure, and the emotional responses of the subjects were measured by the GSR. Significant differences were obtained in the mean responses between death words and non-death

words. These results indicate that the presentation of the death symbol words elicited, at least, an unconscious affective reaction. This evidence seems to support the hypothesis being tested.

To obtain some measure of the extent to which the death concept entered into the subjects' responses, the same 20 death symbol words, which formed a part of the word association list, were used to set up a written test of interassociation. The test was aimed at obtaining an index of the tendency of the subjects to associate these words to the death concept. The results indicate that the tendency to associate these words to the death concept was significantly greater than for any other referent. Although the associative content measure is of an indirect nature, it seems that the emotional responses obtained on the GSR were actually related to the death concept.

The evidence obtained seems to demonstrate the affective nature of the death concept, and is consistent with psychoanalytic theory concerning death symbols. The data are open, however, to other interpretations than the psychoanalytic. Cautious interpretation is called for in any case in view of the indirect nature of the evidence.

REFERENCES

1. BLOCK, V. Nouveaux aspects de la methode dite "psychogalvanique" et son utilisation comme critere de tension affective. *Rev. Neurol.*, 1952, **86**, 168–171.
2. COOMBS, C. H. Adaptation of the galvanic response to auditory stimuli. *J. exp. Psychol.*, 1938, **22**, 244–268.
3. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1954.
4. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
5. IBOR, J. JUAN LOPEZ. Psychiatry and the existential crisis. In F. J. Braceland (Ed.), *Faith, reason and modern psychiatry.* New York: Kenedy, 1955.
6. JONES, H. E. The study of patterns of emotional expression. In M. L. Reymert (Ed.), *Feelings and emotions. The Mooseheart Symposium.* New York: McGraw-Hill, 1950.
7. LINDQUIST, E. F. *Statistical analysis in educational research.* New York: Houghton Mifflin, 1940.
8. MCCLEARY, R. A., & LAZARUS, R. S. Autonomic discrimination without awareness: An interim report. *J. Pers.*, 1949, **18**, 171–179.
9. MCCLEARY, R. A., & LAZARUS, R. S. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.
10. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
11. MCGINNIES, E. Discussion of Howes and Solomon's note on "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 235–240.
12. MERRILL, R. M. The effect of pre-experimental and experimental anxiety on recall efficiency. *J. exp. Psychol.*, 1954, **48**, 167–172.
13. ROSANOFF, A. J. *Manual of psychiatry.* New York: Wiley & Sons, 1927.
14. SIMMEL, E. Self preservation and the death instinct. *Psychoanal. Quart.*, 1944, **13**, 160–185.
15. THORNDIKE, E. L. *The teacher's word book.* New York: Columbia Univer., 1927.
16. THOULESS, R. H. The technique of experimentation on the psychogalvanic reflex phenomenon and the phenomenon of Tarchanoff. *Brit. J. Psychol.*, 1930, **21**, 219–239; 309–321.
17. VENABLES, P. H. The relationship between PGR scores and temperature and humidity. *Quart. J. Psychol.*, 1955, **7**, 12–18.
18. WIKLER, A. The uses of drugs in psychiatric research. *Amer. J. Psychiat.*, 1956, **112**, 961–969.
19. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

Received January 29, 1957.

THE SPECIFICITY OF INTOLERANCE OF AMBIGUITY MEASURES[1]

DOUGLAS T. KENNY AND ROSE GINSBERG[2]
University of British Columbia

Much of the research stimulated by *The Authoritarian Personality* (1) tacitly assumes that intolerance of ambiguity is a generalized trait. At the present time, however, the basic data to substantiate or vitiate such an assumption are surprisingly not available, in spite of the plethora of measures purporting to measure intolerance of ambiguity and the fact that Frenkel-Brunswik (10, 11) expressed caution about assuming the generality of the construct when she initially introduced it. As far back as 1949 Frenkel-Brunswik, noting that the evidence for its generality was only suggestive, stated "... a much wider array of both techniques and population samples would be necessary to establish this generality with an adequate degree of definiteness" (11, p. 131). While the array of techniques for assessing intolerance of ambiguity has increased, there unfortunately has been no corresponding demonstration that the various measures intercorrelate.

The present study is concerned with examining the construct validity (3) of proposed measures of intolerance of ambiguity in terms of two procedures. The first method tests the hypothesis that a positive relationship exists between various tests presumed to measure intolerance of ambiguity. The second approach investigates the hypothesis that proposed measures of intolerance of ambiguity are positively related to authoritarianism. In light of the multiplicity of tests offered by various investigators as measures of intolerance of ambiguity, it was decided to sample a wide range of techniques. Positive intercorrelations would support the validity of the measures as indices of intolerance of ambiguity as a generalized trait. Failure to obtain significant correlations among the measures would indicate either that intolerance of ambiguity is not a unitary characteristic or that the measures are not relevant indicators of the construct.

Particular attention was given to the relevancy of the tests to the construct. In the following, a test is judged to be relevant to the construct of intolerance of ambiguity if it (*a*) elicits behavior specifically mentioned by Frenkel-Brunswik as relevant to her construct or (*b*) had been used by previous investigators as a measure of intolerance of ambiguity. With respect to the first criterion, Frenkel-Brunswik (10, 11, 12) has set forth many behavioral signs in detail, indicating that this disposition is manifested behaviorally in resistance to reversal of apparent fluctuating stimuli, the early selection and maintenance of one solution in a perceptually ambiguous situation, inability to allow for the possibility of good and bad traits in the same person, acceptance of attitude statements representing a rigid, white-black view to life, seeking for certainty, a rigid dichotomizing into fixed categories, premature closure, and remaining close to familiar characteristics of stimuli. In regard to the second criterion, a measure is regarded as having provisional relevancy if Frenkel-Brunswik or some other investigator has published a test purporting to measure intolerance of ambiguity. Of the twelve measures all are analogous to or slightly modified versions of those used by previous investigators. Three are analogous to Frenkel-Brunswik's reversible figure task (11). Detailed considerations of relevancy leading to the selection of each test are given in a later section.

In addition to the evidence provided by the intertest correlations between purported measures of intolerance of ambiguity, this study evaluates their construct validity by testing the hypothesis, derived from the theoretical network defining the construct, that intolerance of ambiguity and authoritarianism are positively correlated.

Previous research on the relation between intolerance of ambiguity and authoritarianism has produced conflicting results. Following Frenkel-Brunswik's provocative but tentative and exploratory findings (9, 10, 11, 12), a number of investigators (2, 8, 16, 18, 19, 20) have obtained evidence for a positive relation-

---

[1] This study was supported by a research grant, A.P. 25, from the National Research Council, Canada.

[2] Now at Indiana University.

ship. However, considerable contradictory evidence (4, 5, 6, 7, 15) has also been reported. Coulter's finding, as reported by Eysenck (8), that a questionnaire measure of intolerance of ambiguity and a test involving a series of pictures of a dog gradually changing into a cat correlate positively, is the only available evidence of a positive relationship between various tests of intolerance of ambiguity.

METHOD

Test Situations

The test situations were administered individually to the *S*s. The experimenter (R. G.) adopted a friendly and permissive approach to the *S*s who were not informed of the purpose of the study or of the true nature of the tasks.

The test battery consisted of twelve measures of intolerance of ambiguity, one attitude measure of authoritarian-submission, and seven measures of aggression. The data derived from the aggression measures will be reported elsewhere. In order to facilitate reading of the correlational data, scoring of some variables was reflected so that a high score always signifies a high degree of intolerance of ambiguity.

Reversal fluctuations. Frenkel-Brunswik (11, p. 127) has postulated that intolerance for a high rate of fluctuation in reversible figures is indicative of intolerance of ambiguity. Such tasks meet the second criterion of relevancy in that they have previously been used by Jones (13). The Necker cube, Beaunis' cubes and Schroder's staircase were presented under three conditions of instructions and three scores were obtained. Preliminary instructions, along with a practice figure, were given in order to acquaint *S* with the nature of reversals. The figures were presented for a time of one minute per figure, each figure being presented under all three conditions before proceeding to the next figure. The three instructional sets were as follows:

1. The *passive set* instructions required *S* simply to watch the figure and report each time she witnessed a change. After *S* reported the first reversal, she was asked to describe exactly what she had seen in order to avoid the continuance of any imaginary changes. The score was the number of reversals occurring in the sixty-second interval, with the score being the total of all reversals reported for all three figures. This scoring was reflected so that a high score would indicate intolerance of ambiguity.
2. Under *active set* conditions, *S* was instructed to "try to make the figure change as often as possible." The active set score was the total number of reversals during sixty seconds per figure, reported for all three figures. This variable was also reflected.
3. The *arrested set* instructions emphasized that *S* should try to stop the figure from reversing and to keep it fixed in one position. The score comprised the total reversals for all three figures, and it was reflected.

Walk's Ambiguity Intolerance Scale.[3] This attitude measure of intolerance which O'Conner (16) found to correlate positively ($r = .55$) with the California E scale consists of eight statements. An illustrative item is: "Nobody can have feelings of love and hate toward the same person." Four response categories were provided for each item, ranging from Agree, 4, to Disagree, 1, with two undecided categories, namely, Undecided: Probably Agree, 3, and Undecided: Probably Disagree, 2. A high score on the A scale is considered indicative of intolerance of ambiguity. The rationale for this scale meets both criteria of relevancy. Frenkel-Brunswik has indicated that intolerant individuals will endorse statements "... expressly designed to reveal a dichotomizing attitude, a rejection of the different, or an avoidance of ambiguities in general" (11, p. 123). The items on this scale express such dichotomizing tendencies.

Trait Discrepancy Scale.[4] According to Frenkel-Brunswik, the person who is tolerant of ambiguities will be able to tolerate and recognize the "... actual coexistence of positive and negative features in the same object. . ." (11, p. 115). Working on the assumption that intolerant persons would not be able to accept the coexistence of an undesirable personality trait and a socially desirable trait in the same individual, Steiner (19) devised a trait discrepancy scale that provided him with a measure of intolerance correlating positively with ethnocentrism. Using Steiner's method of scale construction, a similar scale was constructed, 25 self-concepts being rated for social desirability on a 7-point rating scale by 67 university undergraduates (28 females and 39 males). The judges were instructed to rate each trait on a continuum of social desirability and the median rating for each trait served as its social desirability scale value. The trait discrepancy scale was then constructed by pairing two sets of traits together, one of the sets having similar scale values with respect to social desirability and the other widely discrepant values. The final scale consisted of 15 such pairs.

In the test situation the scale was presented to *S* who was instructed to cross out the pair of characteristics less likely to occur together in the same person. For example, item seven was: (obedient-economical) (obedient-leisurely). The traits judged to be discrepant in social desirability scale values are "obedient-leisurely." The score was the number of discrepant traits crossed out.

Interpersonal Kind-of-Person Situation. In this test, as used by Martin (14) *S* is instructed: "Now I would like you to describe what kind of person you are." Two minutes were allowed for the task and the score was the

---

[3] A copy of this scale, together with the 12 attitude items measuring authoritarian submission employed in deriving the Undecided Answers measure, and the Trait Discrepancy scale has been deposited with the American Documentation Institute. Order Document No. 5462 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[4] See Footnote 3.

Bureau Ednl. Research
DAVID HA... COLLEGE
Dated

number of questions asked by S during that period. In response to any questions from S, the E replied with a set of standardized answers used by Martin. A high score is assumed to indicate intolerance of ambiguity on the grounds that "E confronted the S with an indefinite and ambiguous task and gave vague initial instructions. There were actually no correct or right responses in the situation, only a variety of alternative responses. The S's behavior in regard to clarifying the ambiguity was quantified by counting the number of questions that the S asked before he was willing to impose his own interpretation or structure upon the task and proceed accordingly (14, p. 495)."

Interpersonal Block Situation. In this test Martin's procedure and instructions (14) were followed exactly but with some modifications of the materials. Whereas Martin presented an assortment of Vigotsky blocks to S, the present adaptation consists of presenting to S a large square of cardboard on which five design blocks from the Kohs Test and two sets of the Goldstein-Scheerer Color Form blocks are randomly assigned. S was instructed as follows: "I would like you to figure out what is to be done with these and then go ahead and do it." Three minutes were allowed for the task and the score was the number of questions asked during that time. As in the previous test, Martin's standard procedure for answering S's questions was followed. As with the preceding test, S is confronted with an ambiguous situation with the possibility of many different responses; the S who is intolerant of ambiguity is expected to ask many questions. Martin found significant correlations between his interpersonal measures and the aniseikonic illusion.

Total Questions. The number of questions asked during the one and a half hour experimental session (excluding the questions asked during the two preceding tasks) was the score for this measure. As the whole battery of tests presented a somewhat ambiguous situation to S in terms of the purpose of the experiment and the nature of the experimental tasks, it was assumed that an S who asked many questions in this connection or posed specific questions about the procedures was attempting to structure the ambiguous situation. It will be recalled that E did not explain the nature or the purpose of the experiment. The E had merely stated that most psychological experiments necessitate the omission of such explanations. The E's approach to the Ss was permissive and nonauthoritative in order that any submissive tendencies that the intolerant individuals might have to authority figures might be overcome.

Autokinetic Effect. Block and Block (2) and Taft (20) report evidence that the intolerant individual stabilizes his judgment more quickly than others in the autokinetic experiment. Accordingly, the autokinetic situation was included in this study. Because the number of cues is minimal the autokinetic experiment would seem to present an S with a markedly ambiguous situation.

The apparatus used was a small container, 4¼″ in diameter and 4¼″ in height, in which a 15-watt bulb was inserted. A small aperture, about 1 mm. in size, was drilled in the side of the container and covered with four layers of thin red tissue. To heighten the autokinetic effect, a noisy motor was connected to a timer which also controlled the light, both being activated for 15 seconds during each of 10 trials.

To prepare the S for the autokinetic effect, detailed instructions were given prior to entering the apparatus room. In brief, the instructions emphasized that E would turn out the light when they entered a dimly illuminated room and that when E said "now" a red pin point of light would appear directly in front of the S at about eye level for 15 seconds. After the light went out, S was to estimate three things: the distance the light moved, the direction in which it moved and how confident she was about her judgment of direction. The S was told that she should not worry too much about the accuracy of her judgments of distance and direction, as such estimates were difficult to make under these conditions. When S fully understood the nature of the task, the apparatus room was entered. The S was seated in front of the apparatus at a distance of 12 feet. The autokinetic light source could not be seen by S from where she was seated, nor had the distance of this light source from the S's chair been mentioned. When S was seated, E immediately darkened the room and S was adapted in darkness for about four minutes before the 10 trials began. During this period, E again reviewed the general instructions. The obtained measures were:

1. The autokinetic *mean discrepancy measure* was obtained by grouping the ten trials in successive blocks of two trials, starting from the first trial. The mean movement for each of the five blocks was calculated. If we let $i = 1, 2, 3, 4, 5$ represent the five successive blocks, an S's mean discrepancy score was $X_i = (\bar{X}_1 - \bar{X}_2) + (\bar{X}_2 - \bar{X}_3) + (\bar{X}_3 - \bar{X}_4) + (\bar{X}_4 - \bar{X}_5)/4$. This variable was reflected in order that a high score would indicate intolerance of ambiguity. In keeping with Frenkel-Brunswik's (11, p. 127) assumption that intolerance of ambiguity is shown in the tendency to fix upon one of many response possibilities in an equivocal situation, persons intolerant of the unstructured autokinetic situation should be prompt in establishing an anchoring point.

2. A *consistency score* was obtained by assigning a score of one for each movement consistent with the direction of movement of the preceding trial. According to Frenkel-Brunswik's formulation (11, 12) that the

TABLE 1
CHARACTERISTICS OF THE MEASURES OF INTOLERANCE OF AMBIGUITY AND AUTHORITARIANISM

Test	Mean	*SD*	Range
Passive, Reversals	13.79	7.61	2–50
Active, Reversals	28.49	17.28	6–94
Arrested, Reversals	11.47	6.02	0–28
Walk's A Scale	17.09	3.62	8–25
Trait Discrepancy	7.37	1.95	2–12
Kind-of-Person	.88	.89	0–7
Blocks	1.67	1.27	0–6
Total Questions	3.79	3.07	0–16
Discrepancy, Autokinetic	3.41	3.98	0–26.5
Consistency, Autokinetic	3.82	2.88	0–9
Confidence, Autokinetic	42.11	7.03	22–50
Undecided Answers	5.41	3.58	0–14
Authoritarian Submission	27.54	7.71	12–43

TABLE 2

INTERRELATIONS AMONG MEASURES OF INTOLERANCE OF AMBIGUITY

Test Measures	2	3	4	5	6	7	8	9	10	11	12	13
1. Passive, Reversals	.34**	.31**	−.08	−.11	−.02	−.12	−.05	.22*	−.15	.13	−.01	.01
2. Active, Reversals		.14	.00	−.17	−.03	−.06	−.03	−.18	−.19	.01	−.08	.07
3. Arrested, Reversals			.08	.01	.02	.06	−.16	−.17	.05	−.01	.16	.20
4. Walk's A Scale				.06	−.09	.32**	.01	−.02	−.19	−.10	−.11	.57**
5. Trait Discrepancy					.17	.07	.16	−.01	.02	.06	−.08	−.07
6. Kind of Person						.19	.01	.01	−.08	−.03	−.07	.03
7. Blocks							.38**	.07	.02	.04	−.09	.07
8. Total Questions								−.20	−.01	.00	−.24*	.25**
9. Discrepancy, Autokinetic									.05	−.18	−.03	−.03
10. Consistency, Autokinetic										.17	.03	.14
11. Confidence, Autokinetic											−.26*	−.01
12. Undecided Answers												.17
13. Authoritarian-Submission												

* $p < .05$, two-tailed test.
** $p < .01$, two-tailed test.

individual intolerant of ambiguous situations tends to be rigid and consistent in his behavior when placed in ambiguous situations, the person intolerant of ambiguity should be consistent in the direction she saw the light move.

3. A *confidence estimate* was obtained for each trial, as already mentioned, in terms of very uncertain, somewhat uncertain, average amount of certainty, somewhat confident and very confident, scored from 1 to 5 respectively. The final confidence estimate score was the summation of the rating scores for all the trials. It is assumed that the intolerant person should obtain a high score, in the attempt "to deny external ambiguity as long as such denial can be maintained" (11, p. 129).

Undecided Answers. This score was derived from the answers to the eight items on the Walk scale and the 12 attitude items measuring authoritarian-submissive tendencies, to be described next. It will be recalled that the response categories to the Walk scale provided two "undecided" classifications. The authoritarian-submissive scale was similar in this respect. It is assumed that an *S* who checked either of the two undecided categories displayed tolerance for ambiguity. As the total score was the number of undecided response categories checked, this variable was reflected in the correlational study. McCandless and Holloway (15) made a similar assumption about extreme sociometric ratings. The rationale for relevancy rests upon Frenkel-Brunswik's formulation that the person intolerant of ambiguity expresses his feelings and reactions "in terms of the ends of a continuum rather than a continuum proper" (12, p. 394).

Authoritarian Submission Attitude Scale.[5] The results of the factor analysis of authoritarian items by O'Neil and Levinson (17) provided the basis for the selection of items in this measure. The scale consists of 12 items which had a factor loading of .35 or higher on their Authoritarian-Submission factor. Of the 12 items, five were from the original F scale, one was from the E scale and six were from a Traditional Family Ideology Scale. The response categories and the method of scoring were identical to those used with the Walk scale.

Subjects

The *S*s were 76 volunteer female adults. Their mean age was 23.84 years with a standard deviation of 7.47. Their education ranged from the eleventh grade to graduation from university, with an average of 13.43 years of education. Fifty-six of the *S*s were drawn from the clerical staff of the University of British Columbia. Of the remaining *S*s, 15 were teachers, 4 were students and one was a nurse.

RESULTS AND DISCUSSION

Table 1 shows the means, standard deviations and ranges of the various measures. Intercorrelations between the measures are given in Table 2. The most outstanding feature of Table 2 is the lack of evidence for positive relationship between the various measures of intolerance of ambiguity. With a two-tailed test of significance, four of the 66 correlations are significant at the .01 level of significance, and another three correlations are significant at the .05 level of confidence. Even if a one-tailed test had been employed on the grounds that the hypothesis of a positive relationship between measures of intolerance of ambiguity dictates the direction of the correlation, the main trend of the findings would not be altered. Moreover, two of the seven significant correlations obtained are not in the direction to be predicted from theory. It is also clear from Table 2 that, in the main, the purported measures of intolerance of ambiguity do not correlate significantly with the authoritarian-submission scale, indicating that the hypothesis of a positive relationship between authoritarian-submission and intolerance of ambiguity is not sustained.

---

[5] See Footnote 3.

Only two of the 12 correlations employed to test this hypothesis are significant.

In so far as the present measures may be regarded as relevant indicators of the intolerance of ambiguity construct, the results offer little support for a general construct of intolerance of ambiguity. It would be unwarranted to infer that a person who is intolerant in one situation will be intolerant in all equivocal situations. Moreover, it would be equally unwarranted to assume that intolerance of ambiguity is invariably linked with authoritarianism. These conclusions do not mean, however, that the construct should necessarily be discarded. It may, rather, be less general or broad in scope than had been initially assumed. Future research may discover a number of distinct or relatively independent dimensions of intolerance of ambiguity rather than just one unique generalized factor.

Another inference from the data is that the nature of the task is a crucial element in the discrepant results of studies in this area. If no common factor of intolerance of ambiguity runs through all the tests employed in its name, then the discordant results are not too surprising. This conjecture seems quite plausible in light of our failure to find any such unique factor.

SUMMARY

The major focus of the present study was a test of the construct validity of proposed measures of intolerance of ambiguity. Seventy-six female *S*s were given a battery of tests purporting to measure intolerance of ambiguity and authoritarian-submission. Only seven of the 66 correlations among measures of intolerance of ambiguity were significant at the .05 level, two of these having a relationship opposite to those predicted. In the main, these measures did not correlate significantly with a scale of authoritarian-submission.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. BLOCK, J., & BLOCK, JEANNE. Intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1950, **19**, 303–311.
3. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.
4. DAVIDS, A. Some personality and intellectual correlates of intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415–420.
5. DAVIDS, A. The influence of ego-involvement on relations between authoritarianism and intolerance of ambiguity. *J. consult. Psychol.*, 1956, **20**, 179–184.
6. DAVIDS, A., & ERIKSEN, C. W. Some social and cultural factors determining relations between authoritarianism and measures of neuroticism. *J. consult. Psychol.*, 1957, **21**, 155–159.
7. ERIKSEN, C. W., & EISENSTEIN, D. Personality rigidity and the Rorschach. *J. Pers.*, 1953, **21**, 386–391.
8. EYSENCK, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.
9. FRENKEL-BRUNSWIK, ELSE. Tolerance toward ambiguity as a personality variable. *Amer. Psychologist.*, 1948, **3**, 268. (Abstract)
10. FRENKEL-BRUNSWIK, ELSE. A study of prejudice in children. *Hum. Relat.*, 1948, **1**, 295–306.
11. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual variable. *J. Pers.*, 1949, **18**, 108–143.
12. FRENKEL-BRUNSWIK, ELSE. Personality theory and perception. In R. R. Blake and G. V. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 356–419.
13. JONES, M. B. Authoritarianism and intolerance of fluctuation. *J. abnorm. soc. Psychol.*, 1956, **50**, 125–126.
14. MARTIN, B. Intolerance of ambiguity in interpersonal and perceptual behavior. *J. Pers.*, 1954, 494–503.
15. MCCANDLESS, B. R., & HOLLOWAY, H. D. Race prejudice and intolerance of ambiguity in children. *J. abnorm. soc. Psychol.*, 1955, **51**, 692–693.
16. O'Conner, P. Ethnocentrism, "intolerance of ambiguity," and abstract reasoning ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 526–530.
17. O'NEIL, W. M., & LEVINSON, D. J. A factorial exploration of authoritarianism and some of its ideological concomitants. *J. Pers.*, 1954, **22**, 449–463.
18. SIEGEL, S. Certain determinants and correlates of authoritarianism. *Genet. Psychol. Monogr.*, 1954, **49**, 187–229.
19. STEINER, I. D. Ethnocentrism and tolerance of trait "inconsistency." *J. abnorm. soc. Psychol*, 1954, **49**, 349–354.
20. TAFT, R. Intolerance of ambuiguity and ethnocentrism. *J. consult. Psychol.*, 1956, **20**, 153–154.

Received May 28, 1957.

PERCEPTUAL REACTIVITY TO THREAT AND NEUTRAL WORDS UNDER VARYING EXPERIMENTAL CONDITIONS[1]

EMORY L. COWEN AND PAUL A. OBRIST

University of Rochester

THE purposes of the present research are twofold: (*a*) to investigate the effect of systematic variation of expectancy on perceptual report thresholds for threat and neutral words and (*b*) to provide an experimental test of the usefulness of "hypothesis theory" in explaining the elevation of thresholds for report of threat words observed in some previous investigations.

In an early controversial study, McGinnies (14) found significantly higher thresholds for report of taboo words when compared to neutral ones. That these data offer support for the existence of some organismic "defense" process, has been challenged by Solomon and Howes (17), among others, who argue that McGinnies failed to control for differential familiarity for the two classes of word stimuli, as well as for deliberate or conscious inhibition of report. Subsequent investigations including those of Eriksen and Brown (4), Carpenter, Wiener, and Carpenter (1), and Cowen and Beier (2) in the area of perception, as well as studies by Lazarus, Eriksen, and Fonda (11) and Lazarus and Longo (12), in which personality attributes or clinical typologies have been related to differential reaction to emotional and nonemotional stimuli, has led to a reaffirmation (3) of the concept of "perceptual defense."

On the other hand, another line of research has developed as exemplified by the work of Freeman (5, 6), Lacy, Lewinger, and Adamson (9), Postman, Bronson, and Gropper (16), and others, which has moved in the direction of rejecting this theoretical position. Instead it has been proposed that the phenotypical phenomena frequently cited as illustrations of "perceptual defense" may be more parsimoniously viewed as special instances of already established perceptual principles such as set, or hypothesis behavior. Illustratively, in the experimental study by Postman *et al.* (16) it was found that by controlling for the differential word frequency of taboo and neutral words, response thresholds for the same taboo words used earlier in the McGinnies study (14) were now significantly *lower* than those for frequency-matched neutral words. Elsewhere Postman proposes:

> What appears to be perceptual defense results from the dominance of strong alternative hypotheses rather than from active repression of the inimical or dangerous. In the presence of partial information, strong hypotheses incompatible with the threatening stimulus may be evoked. . . . If this is the case the subject will appear to be defending himself against perception. If, however, hypotheses related to the negative stimuli are strong, the opposite of defense will appear to operate (15, p. 300).

In the present research, strength of hypothesis for threat and neutral words has been varied systematically. From Postman's theoretical position it would be predicted (*a*) that as expectancy or set for threat words is increased experimentally, the discrepancy in threshold of report for threat and neutral words will decrease, and (*b*) that as 100 per cent expectancy is reached for threat words, thresholds of report for threat words will become (significantly) lower than those for neutral words.

METHOD

Thirty male and thirty female introductory psychology student volunteers served as *S*s in the present research. These *S*s were randomly assigned by sex to one of three experimental groups, each of which contained ten males and ten females.

All *S*s were initially informed that they were to decipher words from a series of booklets consisting of thirty carbon copies of a single five-lettered word, electrically typed in capital letters, and arranged in order from the most blurred copy to the clearest. This technique is described more fully elsewhere (2). Following two practice booklets the three experimental preconditions were initiated. A list of 36 words was read to each *S* with the instruction that the final test series would contain some words similar to those on the reading list. There were three groups, as follows:

Group I (0 per cent threat expectancy) was read only neutral words.

Group II (33 per cent threat expectancy) was read a list containing 12 threat words and 24 neutral words.

Group III (100 per cent threat expectancy) was read all threat words.

---

[1] Portions of the present paper were presented to Division 8 at the Amer. Psychol. Ass. meetings in Chicago, August 1956.

TABLE 1

PAIRED THREAT AND NEUTRAL WORDS WITH FREQUENCIES

Threat	Neutral	Frequency[a]
Bitch	*Juror*	4
Penis	*Rayon*	1
Whore	*Derby*	3
Death	*Clear*	100+
Urine	*Humid*	1
Cheat	*Rural*	18–19
Stink	*Pivot*	4
Flunk	*Sedan*	2

[a] Based on general frequency of occurrence in one million words (18).

The threatening or neutral nature of the words for the reading list (as well as the final test series) had previously been established empirically by means of ratings by graduate students in psychology, with respect to the dimension of stimulus threat value. Only those words with consistent interjudge agreement in designation as either threat or neutral words were included in the reading lists and final test series. For all groups, *S*s were asked to repeat each word after the *E*. This precaution was taken not only to make sure that the *S* had heard the word correctly, but also to give *S*s experience in the actual verbalization of threatening stimuli.

All *S*s were then given the test booklets, in random order, to decipher. This final test series, individually administered, consisted of 16 words, eight threat and eight neutral, *none* of which had appeared on any of the earlier reading lists. The threat words contained four of the "sexually-taboo" variety, and four others that did not involve any sexual taboo. Each of the eight threat words had been matched with a neutral word on the basis of identical tabled frequencies according to the Thorndike-Lorge word frequency count (18). Where no word frequency was given for a threat word, an arbitrary frequency of one was assigned, and matching of the neutral word proceeded on this basis. Table 1 presents the test series of words and their frequencies.

The *S*s were allowed two seconds per page after which they were required to turn to the next page. The *E* recorded total trials required for correct report of each word. Guessing was encouraged, and all prerecognition guesses were recorded. At the end of the perception experiment, all *S*s were asked to write down as many words as they could recall from the original pre-experimental list of words read to them. They were given three minutes for this task.

RESULTS AND DISCUSSION

Report thresholds. Mean thresholds for threat and neutral words as well as the departure of the mean D score between these two classes of words from zero are presented for each group in Table 2. Table 3 summarizes a three-way analysis of variance for perceptual report thresholds, the main effects of which include type of words (threat vs. neutral), sex, and experimental conditions (level of set). Since all *S*s were given both classes of word stimuli, this effect is regarded as a within subject effect. Accordingly, a mixed design Type III, as described by Lindquist (13), was employed as the basic model. Relative to the purposes of the present experiment, the most salient results in this table are the significant words-by-conditions interaction and the significant main effect of words.

The significant words-by-conditions interaction indicates that there are reliable differences in threat minus neutral word thresholds (D scores) for the three experimental conditions. From inspection of the last column of

TABLE 2

MEAN PERCEPTUAL REPORT THRESHOLDS AND D SCORES FOR EXPERIMENTAL CONDITIONS AND SEX

Condition	Sex	Threat	Neutral	D score[a]	Sig. of D score from 0
I (0% threat)	Male	22.15	20.79	+1.36	.05–.1**
	Female	22.15	18.84	+3.31	.01
	Total	22.15	19.82	+2.33	.01
II (33% threat)	Male	21.18	19.93	+1.25	.1–.2
	Female	22.91	20.45	+2.46	.01
	Total	22.04	20.19	+1.86	.01
III (100% threat)	Male	21.21	21.59	−0.38	.5–.6
	Female	21.93	21.08	+0.85	.05–.1
	Total	21.57	21.34	+0.24	.5–.6

[a] Plus (+) signs indicate higher thresholds for threat words.

** *p* value tabled on basis of two-tailed test.

TABLE 3

ANALYSIS OF VARIANCE FOR PERCEPTUAL REPORT THRESHOLDS

(Lindquist Type III Design)

Source	*df*	Mean Squares	*F*	*P*
Between *S*s	(59)			
Sex	1	0.21	—	—
Conditions	2	2.33	—	—
Sex × Conditions	2	11.10	—	—
Error (between)	54	38.42		
Within *S*s	(60)			
Words	1	65.50	28.48	.001
Words × Sex	1	16.05	6.98	.01–.05
Words × Conditions	2	12.11	5.27	.01
Words × Conditions × Sex	2	0.34	—	—
Error (within)	54	2.30		
Total	119			

Table 2, it can be seen that there is a drop in discrepancy between threat and neutral word thresholds, as we strengthen experimentally the set for threat stimuli. Table 2 also suggests that a major component of this drop may be the increase in neutral word thresholds. However, analysis of variance comparing the mean thresholds for neutral words across conditions failed to demonstrate significant differences. We may note that direct analysis of neutral word thresholds compares different *groups* of *S*s, whereas D-score analysis is based on each *S* serving as his own control. In the former instance we thus introduce intragroup variability (error variance) deriving from such sources as differential visual acuity, inconstancies in testing conditions etc., and consequently reduce the likelihood of demonstrating significant neutral word threshold differences. Accordingly, we view the earlier noted decrease in threat-neutral discrepancy from Condition I to Condition III as confirming the first prediction from hypothesis theory, consistent with the earlier findings of Postman *et al.* (16).

The significant main effect of words in Table 3, viewed conjunctively with the actual group means presented in Table 2, indicates that threat words on the whole have significantly higher thresholds than neutral ones. Although it is probable, on the basis of the words-by-conditions interaction, that the several conditions contribute differentially to this main effect, these data are nevertheless in sharp contrast to the empirical findings of Postman *et al.* (16). The latter investigators report that after matching threat and neutral words for word frequency, *lower* report thresholds were found for threat words under *all* of their experimental conditions. Postman *et al.* (16) attribute their results to the probable underestimation of functional word frequencies for threat words in standard word frequency tables. Notwithstanding the rather precise control for word frequency in the present experiment, the Postman *et al.* (16) findings are not duplicated. In fact, with respect to this segment of the analysis, the present results go significantly in the opposite direction.

In addition to this discrepancy at the empirical level it is important to note that the overall higher threat thresholds observed in the present study negate the second prediction deriving from hypothesis theory (i.e., that threat words should be seen significantly more rapidly than neutral ones where a 100 per cent expectancy for the former has been established). Under none of the present experimental conditions, including 100 per cent threat expectancy, are threat words seen more rapidly than neutral ones. In fact, for the 100 per cent threat condition, the overall trend remains in the direction of higher thresholds for threat words, albeit not significantly so.

To this point, the discussion has been limited to the derivations from hypothesis theory, which we have set out to study experimentally. Our findings, however, have some implications for the word-frequency control issue originally developed by Solomon and Howes (17). For the specific experimental conditions and word stimuli used in this study, it is clear that an explanation of the elevation of threat word thresholds in terms of differential word frequency would not be defensible. This is not to say that word frequency cannot or has not constituted an uncontrolled element spuriously influencing thresholds in other experiments. What remains, however, is the fact that threat word thresholds continue to be elevated in a situation where word frequency is either ruled out as a contaminating influence, or if not, is as Postman *et al.* (16) suggest, operating in a direction opposite to the findings (i.e. to lower thresholds for threat stimuli by underestimation of functional usage of threat words).

The present investigators were nevertheless somewhat perplexed in trying to develop a frame of reference in which the discrepancies between our findings and those of Postman *et al.* (16) with respect to the relative elevation or depression of threat word thresholds might be resolved. To us it seemed that a promising lead might be contained in Lazarus' earlier critique (10) of the Postman *et al.* study, in which it was proposed that tabled *threat* word frequencies do not necessarily correspond with their general familiarity to college students. Lazarus further suggested that "... findings in terms of mean differences in thresholds for threat and neutral words could be made to go in either direction by the appropriate selection of words on the basis of such variables as context, contrast, sequential probabilities, and a host of other factors not directly related to word frequency" (10, p. 397).

As we examined the actual word lists utilized in the two investigations it appeared probable to us, quite apart from the threat word issue raised by Lazarus, that many of the *neutral* words from our study had substantially higher functional usage for college students than did neutral words with comparable tabled frequencies, used by Postman *et al.* (16). A small additional study was undertaken in order to subject this hunch to experimental test.

Ten "neutral" words were selected, five from the Postman *et al.* series and five from our own. These words were chosen largely in terms of our preconceptions regarding their functional frequencies. The words were placed in a fixed random order, and given to 30 judges, all of whom were either staff members or gradu-

TABLE 4

NEUTRAL WORDS FROM TWO EXPERIMENTS, ORDER OF PRESENTATION, FREQUENCY, AND JUDGES' RANKS

Words	Order	Frequency	Mean rank	Rank of $\bar{x}$ Rank
Postman *et al.*				
Lathe	2	2[a]	5.90	6
Lotus	4	2	8.23	8
Tiara	7	1	8.30	9
Capon	8	1	8.93	10
Clove	9	8	6.83	7
Mean		2.8	7.66	8
Cowen & Obrist				
Rayon	1	1	2.53	2
Humid	3	1	1.17	1
Juror	4	4	3.50	3
Pivot	6	4	4.33	4
Derby	10	3	5.17	5
Mean		2.6	3.34	3

[a] Based on Throndike-Lorge (18) general count for frequency of occurrence in one million words.

ate students in the psychology department at the University of Rochester. The judges were asked to rank order the 10 words on the basis of estimated frequency of usage by a mixed (male and female) college population. Mean functional frequency ranks were then computed for all words.

Table 4 presents the two sets of words, together with their order of presentation, frequency of occurrence in one million words according to the Thorndike-Lorge general count (18), and mean judge rankings. We note first that the tabled frequencies for the two sets of words are highly comparable. Were there some correspondence between tabled frequencies and estimated functional usage, it might be anticipated that judges' rankings for the latter dimension would be essentially random across the two sets of words. Inspection of Table 4 indicates that this is far from the case. Estimated functional frequency ranks for our words, run from one to five. There is, in fact, not a single "misplacement" of ranks for the words from the two studies. This global impression is visible as well in the analysis of individual word rankings. Three of the Postman *et al.* words, *tiara, lotus*, and *capon* have 75 per cent or more of their ranks at eight, nine, and ten. On the other hand, our word *humid* has 25 "one" ranks and five twos. While we are willing to concede that this specific ranking may, in part, reflect some climatic preoccupation of Rochesterians, the basic trend of differences for the functional frequencies of the two sets of words seems highly reliable.

The foregoing demonstration seems to establish, at least within the range of frequencies studied, the possibility of gross disparity between the tabled values characteristically used in selecting words for "perceptual defense" studies, and estimated functional usage of words. Moreover, the particular frequency band examined is of some importance, since, due to the usually low tabled frequencies for threat words, neutral control words are most often selected from within this range.

The demonstrated discrepancy between tabled frequencies and estimated functional frequencies is alarming! For not only must we face up to past speculations regarding the probable systematic underestimation of threat word frequencies (10), but we must now add to this the discomforting recognition that within the critical low frequency range there may be no correspondence between estimates of word frequency derived from adult and juvenile books, newspapers and magazines on the one hand and the frequencies inherent in the every day language of the college student. Since "perceptual defense" studies have been executed almost unilaterally on the college sophomore, or some variant thereof, it would seem that the reference population for determining word frequencies should most appropriately be such a group in future work. This will require the empirical establishment of word frequency norms for college students (or whatever other reference group is to be studied perceptual defense-wise). The technique we have reported here does no more than to approximate this end very crudely.

One further disconcerting comment may be appropriate. When a "perceptual defense" study, in which word frequency is to be controlled is undertaken, the investigator characteristically matches neutral word frequencies to available tabled frequencies for threat words. In so doing, he has a very substantial pool of low frequency words to draw from, which, on the basis of present findings, may not necessarily correspond to actual usage frequencies. In such a situation there may develop the unintentional but nevertheless quite serious danger of systematic error in the selection of specific words from a group with identical frequencies, on the basis of unspecified biases and predilections. To the extent that, within this key low frequency range, tabled and functional frequencies do not correspond, we are faced with an important problem of representative and systematic design (7). For if we are ultimately interested in functional frequencies, and *n* sets of equated tabled frequencies represent *n* continua of functional frequencies, our fundamental *N* with respect to functional frequency is one, and generalizations about word frequency may be restricted to the (unknown) functional frequencies of the *specific* set of neutral control words employed.

It is quite possible that the reversal in findings of the two studies with respect to relative elevation or depression of threat words thresholds, is in part explainable in terms of differential functional frequencies. However, even though we have demonstrated that some of our own neutral words have higher functional frequencies than some of those with comparable tabled frequencies used by Postman *et al.*, there are no data from either study contrasting the functional frequencies of threat and neutral words. We can here do no more than speculate along with Lazarus (10) that for the pertinent reference population (i.e. college students) the threat words used may have higher functional frequencies than "matched" neutral control words (for both studies). On the basis of such speculation, we believe that our earlier statements concerning elevation of threat word thresholds may be defensible. At the same time, it would be unrealistic to fail to admit that, at least within the critical low frequency range, frequency "matchings" in studies to date may be more a matter of fantasy than fact. The necessary *desideratum* must be the determination of functional frequencies for both classes of word stimuli. To this may be added the ultimate goal of ascertaining individual frequency differences in word usage (10). These ends may only be achieved empirically.

Sex differences. Returning to Table 3, we note a significant words-by-sex interaction, indicating that there is a difference between sexes in the discrepancy between threat word and neutral word thresholds. Table 2 suggests that this difference is due to the consistent tendency for males to have lower report thresholds than females for threat stimuli. These data are in direct support of the earlier comparable finding by Postman *et al.* (16). Of the several possible interpretations offered by the latter investigators to explain this phenomenon, that of possible differential readiness to report threat words seems best to accomodate the data.

Taboo vs. non-taboo threat words. In order to check further on the possible operation of conscious inhibition, the threat words were divided into taboo threat (bitch, penis, whore, and urine) and non-taboo threat words (death, cheat, stink, flunk). Again, using a Lindquist Type III analysis (13), mean D scores for the four taboo pairs vs. the non-taboo threat words were compared. No significant differences were observed in the D scores for the two types of words, although there was some tendency for the taboo words to have higher D scores from their matched neutral words, than did the non-taboo threat words from theirs ($p = .20$). It may be pertinent to note that *S*s were given experience in verbalizing threat words overtly in Groups II and III during the reading period so as to reduce the possibility of deliberate inhibition. Perhaps the reduction in conscious inhibition deriving from this procedure accounts for our failure to find significant differences between taboo and non-taboo threat thresholds in the present investigation.

Prerecognition guesses. A three-way analysis of variance (Type III) was also computed for number of prerecognition guesses. Although these findings are not tabled the results are quite clear-cut. There are two significant effects observable, the main effects of words ($p < .01$) and conditions ($.01 < p < .05$). Inspection of means indicates that more guessing has occurred for threat words than for neutral ones, and that guessing behavior is highest for Group II (33 per cent threat). The former finding, although partially confounded by the greater number of threat word trials, seems to reflect at least some element of genuine differential guessing behavior for the two classes of word stimuli. Illustratively, in all six subgroups, including males in Group III (where threat word threshold is actually slightly lower than neutral word threshold) there are significantly more prerecognition guesses for the threat words.

The increase in guessing behavior noted in Group II is highly reliable for both sexes and for both types of words. Here we might speculate that the crucial determinant is the ambiguity of the initial reading list, which was the only one of the three to contain words from two clearly differentiable *genera*.

Recall behavior. A simple two-way, conditions-by-sex analysis of variance was computed for total number of words from the pre-experimental reading list recalled by *S*s during the postexperimental period. A highly significant conditions effect ($p < .001$) was observed here, with markedly superior recall by Group III (100 per cent threat). These data are explainable in terms of the relatively finite reference population from which the threat words were drawn.

Implications. In relating our findings to current theoretical controversies, it is apparent that the principle of set or perceptual hypothesis behavior is a sufficient basis for accommodating the observed decrease, from Condition I to Condition III, in threat-minus-neutral discrepancies. On the other hand, the elevation of threat thresholds across the board, in contrast with those for neutral thresholds, constitutes a datum not readily explainable within the framework of hypothesis theory as stated by Postman (15). If we proceed from the assumption that the present neutral control words constitute an appropriate matching for the threat stimuli, then the latter finding would be more readily interpretable by a concept of response interference from threatening stimuli, or differential response hierarchy along the lines proposed by Eriksen (3), Eriksen and Browne (4), Hochberg, Haber and Ryan (8), and others. In that case, it would not only be feasible but necessary to view these two presumably antagonistic theoretical explanations as complimentary in order to explain adequately all of the findings of the present study.

If, however, our supplementary findings suggesting a lack of correspondence between tabled and functional frequencies are considered, we must recognize the strong possibility that earlier tests of the relationship between word-type (threat and neutral) and response threshold have been inappropriate.

SUMMARY

In order to test two predictions deriving from hypothesis theory, sixty *S*s divided into three groups were tested for perceptual reactivity to threat and neutral words matched for word frequency. The experimental groups differed only in terms of induced strength of expectation for threat words from 0 per cent to 100 per cent.

The findings support the prediction that as strength of set for threat stimuli increases, the discrepancy between threat and neutral word thresholds will decrease. But contrary to predictions from hypothesis theory, threat words were found to have higher report thresholds than neutral words in all conditions, even with word familiarity controlled.

A supplementary study, however, demonstrated gross disparities between tabled frequencies and estimated functional frequencies of neutral words, giving rise to several fundamental methodological considerations regarding "perceptual defense" research.

Several empirical analyses dealing with prerecognition guessing behavior, and postexperimental recall behavior are reported.

REFERENCES

1. CARPENTER, B., WIENER, M., & CARPENTER, JANETH T. Predictability of perceptual defense behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 380–383.
2. COWEN, E. L., & BEIER, E. G. Threat-expectancy, word frequencies, and perceptual prerecognition hypotheses. *J. abnorm. soc. Psychol.*, 1954, **49**, 178–182.
3. ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175–182.
4. ERIKSEN, C. W., & BROWNE, C. T. Experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, **52**, 224–230.
5. FREEMAN, J. T. Set or perceptual defense? *J. exp. Psychol.*, 1954, **48**, 283–288.
6. FREEMAN, J. T. Set versus perceptual defense: A confirmation. *J. abnorm. soc. Psychol.*, 1955, **51**, 710–712.
7. HAMMOND, K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, **51**, 150–159.
8. HOCHBERG, J. E., HABER, S. L., & RYAN, T. A. "Perceptual defense" as an interference phenomenon. *Percept. Motor Skills*, 1955, **5**, 15–17.
9. LACY, O. W., LEWINGER, N., & ADAMSON, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–174.
10. LAZARUS, R. S. Is there a mechanism of perceptual defense? A reply to Postman, Bronson, and Gropper. *J. abnorm. soc. Psychol.*, 1954, **49**, 396–398.
11. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–483.
12. LAZARUS, R. S., & LONGO, N. The consistency of psychological defenses against threat. *J. abnorm. soc. Psychol.*, 1953, **48**, 495–499.
13. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
14. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
15. POSTMAN, L. On the problem of perceptual defense. *Psychol. Rev.*, 1953, **60**, 298–306.
16. POSTMAN, L., BRONSON, WANDA C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–225.
17. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–271.
18. THORNDIKE, E. L., & LORGE, I. *The teachers' word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., 1944.

Received May 2, 1957.

AUTHORITARIANISM, PUNITIVENESS, AND PERCEIVED SOCIAL STATUS[1]

ALAN H. ROBERTS AND RICHARD JESSOR[2]

University of Colorado

THE problem dealt with in this study is the effect of perceived social status upon the relationship between authoritarianism and punitiveness. Previous studies concerned with this problem (1, 2, 3, 4, 5, 8) have tended to yield somewhat inconsistent results. These inconsistencies may perhaps be resolved in part if the social cues present in the interpersonal situation are taken into account, the relationship between authoritarianism and punitiveness being conceived to be mediated by social stimulus variables whose meaning derives largely from past learnings. Specifically, high authoritarians may be assumed to have learned to avoid expressing overt and direct aggressiveness towards individuals perceived to be of higher status than themselves and instead, to displace this aggressiveness to out-group members, violators of conventional values, and low-status figures in general. Theoretical support for this assumption may be inferred from the descriptions of "authoritarian submission" and "authoritarian aggression" in *The Authoritarian Personality* (1, pp. 230–234).

Four specific hypotheses were investigated.

1. When the perceived status of the frustrators is not considered in the analysis, there is no difference between high and low authoritarians with respect to punitiveness.
2. High authoritarians, more often than low authoritarians, tend to respond with direct hostility and aggressiveness when frustrated by low status figures.
3. High authoritarians, more often than low authoritarians, tend to respond with indirect, displaced hostility and aggressiveness when frustrated by high status figures.
4. With respect to punitiveness, low authoritarians manifest no difference in reaction to high and low status frustrators.

METHOD

Subjects

A scale[3] consisting of 28 items from the California F (authoritarianism) scale (1) was administered to approximately 800 students in an introductory class in psychology at the University of Colorado. From these 800 respondents, the 60 individuals scoring highest and the 60 individuals scoring lowest on the F scale were asked to volunteer for this study. Of these, 48 high scorers and 53 low scorers participated in the experiment. The two groups of *S*s were made up of approximately equal numbers of males and females, and the median age of *S*s in both groups was 20. The mean F-scale score of the High group was 96.75; the *SD* was 5.80. The mean F-scale score of the Low group was 47.00; the *SD* was 5.02.

Stimulus Materials

A series of 24 separate stimulus cards was constructed, each card portraying two individuals, one of whom is shown to be frustrating the other.[4] The cards are very similar in appearance to the *Rosenzweig Picture-Frustration Study* (6), and may be considered as a revision of that technique for our specific purposes. On twelve of the cards in the present study (the "high status cards"), the frustrator is depicted as a person of high social status with respect to the person being frustrated. On the remaining twelve cards (the "low status cards"), the frustrator is depicted as a person of low social status with respect to the person being frustrated. The social status of the stimulus figures was intended to vary along the dimensions of age, education, income, occupation, and recognizable stereotype of authority in our society. Cues appropriate to these dimensions were manipulated in the drawings.

---

[1] This study was supported in part by financial aid from the psychology department of the University of Colorado and from the Office of Naval Research grant to J. R. Gibb.

[2] The first author is now at Elgin State Hospital, Elgin, Illinois. The second author was a Social Science Research Council fellow at the time this article was written.

[3] The scale was of the usual Likert type, but the number of response categories was reduced from six to four to permit machine scoring of the completed questionnaires. Each item was scored 1, 2, 4, or 5 according to whether the *S* disagreed very much, disagreed a little, agreed a little, or agreed very much, respectively. If an item was not answered it received a score of 3.

[4] Reproductions of the 24 stimulus cards used in this study have been deposited with the American Documentation Institute. Order Document No. 5460 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $2.25 for microfilm or $5.00 for photocopies. Make check payable to Chief, Photoduplication Service, Library of Congress.

An example of a high status card is Card 10 (Army Officer to Enlisted Soldier: "I know I promised you a pass this weekend, but you'll have to stay here and stand guard duty"). An example of a low status card is Card 11 (Secretary to Boss: "I'm too busy to type that now").

To determine how successfully the social cues were manipulated in the drawings, the face validity of the stimulus cards was ascertained. After the main part of the experiment was concluded, all *S*s were asked, as part of a "test of perceptual sensitivity," to guess the age, occupation, yearly salary, and years of education of both figures on six of the high and six of the low status cards, exposed for 30 seconds each. These data clearly indicate that the relative status relationships were generally perceived in the intended direction by a large majority of the *S*s.[5] There were no differences between high and low authoritarians in these judgments.

Administration and Scoring

The *S*s were tested in two intermixed groups of approximately 50 *S*s each. By means of an opaque projector, the stimulus cards were shown separately on a large screen for 60 seconds, and *S*s were instructed to write in their answer books the very first reply that came to mind in answer to the comment of the frustrating figure.

Four scoring variables[6] to characterize the responses of the *S*s in terms of punitiveness were defined in detail and denoted by specific examples. The definitions correspond closely to those of Rosenzweig (7). Very briefly, the four variables were:

1. Personal Extrapunitiveness (P)—the response is directly hostile or aggressive toward the frustrator.
2. Nonpersonal Extrapunitiveness (N)—the hostile or aggressive response is displaced or directed toward some person, object, or event other than the frustrator.

The sum of the two foregoing scores was termed Total Extrapunitiveness (E).

3. Intrapunitiveness (I)—blame or censure is directed by *S* upon himself.
4. Impunitiveness (M)—the frustrating situation is construed not to be frustrating or to be beneficial.

Interscorer reliability was assessed by having three psychologists[7] independently score the same 15 protocols. For 360 scored responses, agreement between A and B was 77%, between A and C was 71%, and between B and C was 94%. All protocols were then scored by the first author without knowledge of the *S*s' F-scale score. The total frequency of P, N, I and M scores was computed for each protocol, and the E score was obtained by summing the P and N scores.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF FIVE PUNITIVENESS CATEGORIES FOR THE TOTAL TEST AND FOR HIGH AND LOW STATUS CARDS SEPARATELY

Measure	High F ($N = 48$)		Low F ($N = 53$)		t
	Mean	*SD*	Mean	*SD*	
Total Test:					
E	18.06	2.37	17.57	2.69	.95
P	5.83	2.34	5.14	2.85	1.32
N	12.23	2.29	12.43	2.75	.39
I	1.67	.94	2.00	1.26	1.60
M	4.25	2.19	4.34	2.24	.29
High Status Cards:					
E	9.48	1.47	9.67	1.78	.42
P	.67	1.00	2.75	2.02	6.61**
N	8.81	1.40	6.92	2.08	5.33**
I	.46	.58	.53	.72	.55
M	2.04	1.40	1.77	1.48	.94
Low Status Cards:					
E	8.59	1.46	7.91	1.48	2.25*
P	5.17	1.83	2.39	1.43	8.37**
N	3.42	1.57	5.52	1.40	7.01**
I	1.21	.66	1.47	.80	1.82
M	2.21	1.27	2.57	1.31	1.39

* Significant beyond the .05 level of confidence (two-tailed test).

** Significant beyond the .01 level of confidence (two-tailed test).

RESULTS AND DISCUSSION

The significance of the mean differences between high and low authoritarians with respect to E, I, M, N, and P on all 24 stimulus cards together, and on the 12 high status and 12 low status cards separately, was tested by means of the t test for independent large samples. The results of these comparisons may be seen in Table 1.[8]

With respect to Hypothesis 1, no significant differences in E, I, M, N, or P scores on the test *as a whole* could be found. While negative findings cannot verify a hypothesis, this lack of difference is fully consistent with our first hypothesis.

When the high and low status cards are analyzed separately, the differences between the High and Low F groups with respect to P and N scores are all significant in the predicted direction beyond the .01 level of confidence. We may therefore reject the null statement of Hypotheses 2 and 3. It appears clear that high authoritarians, more often than low authoritarians, tend to respond with personal hostility to low status figures and with displaced hostility to high status figures. Conversely, high authoritarians, less often than low authoritarians, tend to respond with personal hostility to high status figures and with displaced hostility to low status figures. These results suggest that findings of a relationship between total extrapunitiveness and authoritarianism (2, 5, 8) may be at least partly accounted for in terms of this previously overlooked status variable rather than by assuming greater generalized extrapunitiveness on the part of high authoritarians. This interpretation is further supported by the finding in Table 1 of a significant difference between high and low authoritarians with respect to total E on the low status cards taken alone. It is only on these cards that high authoritarians are more extrapunitive than low authoritarians.

---

[5] For example, on Card 3, the stimulus figure depicted as having the higher status was rated to be younger than the low status figure by only 4 of the 101 *S*s, was rated to have less education by 1 of the 101 *S*s, and was rated to have less income by none of the *S*s. The high status figure was perceived as having a business or professional occupation by 96 *S*s, a semiprofessional occupation by 4 *S*s, and was labeled a "gentleman" by one *S*. Similarly, the low status figure was perceived as a waitress by 98 *S*s, a maid by 2 *S*s, and a housewife by the remaining *S*. A complete summary table of the face validity data has been deposited with ADI, cf. Footnote 4.

[6] Complete definitions of the scoring variables have been deposited with ADI, cf. Footnote 4.

[7] The authors are indebted to James Selkin for his assistance in establishing the interscorer reliability of the punitiveness categories.

[8] All raw score data have been deposited with ADI, cf. Footnote 4.

The final hypothesis was to the effect that while the responses of high authoritarians to frustration are, at least in part, mediated by the perceived status of the frustrator, low authoritarians tend to respond to frustration relatively independently of the status of the frustrator.

In order to test Hypothesis 4, the difference between the mean P score on the high status cards and on the low status cards was statistically evaluated for the High and for the Low F groups separately. In the High F group this difference was clearly significant beyond the .01 confidence level (t, for related measures, of 17.31). The same analysis for the Low F group showed the difference to be insignificant. These data, therefore, strongly support Hypothesis 4; high authoritarians respond with direct personal hostility to low status figures significantly more often than they do to high status figures. Low authoritarians appear to respond with direct personal hostility independently of the status of the frustrator.

Hypothesis 4 includes a prediction about displaced or indirect hostility as well as about direct aggression. Hence, the same kind of analysis as described for P scores was made for the N score. This analysis shows that *both* high and low authoritarians tend to respond with displaced hostility to high status figures more often than they do to low status figures, both cases significant beyond the .01 level of confidence. This result does not accord with Hypothesis 4, but it is somewhat tempered by the fact that the difference for the High F group (5.40) is significantly greater than the difference for the Low F group (1.40), significant beyond the .01 level of confidence. In short, the data tend to support Hypothesis 4, although one finding does require certain reservations.

Several limitations of this study must be noted. In the first place, the experimental groups included only the most extreme segments of an initial group of 800 *S*s. Generalization to the larger group of less extreme cases is not warranted on the basis of this study alone. Second, the semiprojective test used in this research was not fully refined. Some of the stimulus cards proved to be undiscriminating, and the technique as a whole would undoubtedly benefit from further refinement involving item analysis procedures. Finally, it should be noted that all the hypotheses of the present study were tested by semiprojective test data rather than by overt behavior in life situations. The findings are, therefore, limited to whatever extent these two classes of behavior do not directly covary. It is obvious that perfect covariation of test behavior and overt behavior is unlikely, especially where punishment frequently follows the responses in life situations. Nevertheless, the present hypotheses would be maintained—and are testable—in behavioral situations where factors likely to inhibit free responding are controlled or minimized.

Our orientation to this research was that, in predicting the behavior of individuals on the basis of presumed personality dimensions such as authoritarianism, no general conclusions should be drawn without taking into account the specific situational cues that may mediate this behavior. Within the limitations of this study, we have shown that punitiveness is not a general tendency that may be identified with authoritarianism, but rather that the punitive behavior of an individual in an interactional situation is dependent upon both his authoritarianism and certain social status cues present in his stimulus environment.

SUMMARY AND CONCLUSIONS

Some of the interrelationships among authoritarianism, punitiveness, and status were studied by means of a semi-projective technique based on the *Rosenzweig Picture-Frustration Study*. It was found that when punitiveness is measured by this technique, high and low authoritarians do not differ with respect to extrapunitiveness independently of status. When social status is considered, however, high authoritarians, more often than low authoritarians, tend to respond with personal hostility to frustrators of low status and with indirect or displaced hostility to frustrators of high status. The data with respect to our hypothesis that low authoritarians respond to frustrators independently of their status are somewhat ambiguous but generally affirmative. We conclude that punitiveness is a function of both the personality of the individual and the particular stimulus environment that elicits the behavior.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. GUBA, E. G., & GETZELS, J. W. Frustration reaction and authoritarianism. *Amer. Psychologist*, 1956, 362. (Abstract)
3. LINDZEY, G. An experimental test of the validity of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, **18**, 315–320.
4. LINDZEY, G. Differences between the high and low in prejudice and their implications for a theory of prejudice. *J. Pers.*, 1951, **19**, 16–40.
5. ROBERTS, A. H., & JOHNSON, D. M. The relationship of ethnocentrism to punitiveness as measured by the Rosenzweig Picture-Frustration Study. Paper read at Midwest. Psychol. Ass., Cleveland, 1952.
6. ROSENZWEIG, S. The picture association method and its application in a study of reactions to frustration. *J. Pers.*, 1945, **14**, 3–23.
7. ROSENZWEIG, S., FLEMING, E., & CLARKE, H. Revised scoring manual for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1947, **24**, 165–208.
8. WILLIAMS, H. H. An investigation of extrapunitiveness as measured by the Rosenzweig Picture-Frustration Study and the California scale of the antidemocratic personality. Unpublished master's thesis, Univer. of Pittsburgh, 1949.

Received May 10, 1957.

CONFORMITY AND PROBLEM SOLVING[1]

CHARLES Y. NAKAMURA

University of California, Los Angeles

RECENT studies have expressed considerable interest in the relation between problem solving and various nonintellectual personality variables. Impetus in this direction has been provided, in part, by the fact that in problem-solving tasks individual differences are repeatedly found that are not accounted for by differences in general intelligence as measured by the conventional tests. In the search for factors that contribute to this variability, a number of experimenters have investigated nonintellectual variables (6, 12, 13). Although by no means conclusive, these studies present considerable evidence that such variables do contribute to individual differences in problem-solving performance.

This experiment was undertaken to investigate the possibility that conformity, a nonintellectual variable, may be shown to account for variation in problem-solving scores that is not accounted for by scores on a conventional intelligence test. In this study, conformity refers to behavior that is induced by subjecting individuals in a laboratory setting to conflict-producing social pressure. The experimental situation is a modification of that originated by Asch (2). The typical Asch situation is one of conflict for the naive subject who must either respond in accord with his own perception, thus disagreeing with the majority opinion, or respond in agreement with the group contrary to his own perception. An individual is said to conform if he resolves the conflict by agreeing with the contrived group consensus. This concept of conformity is to be distinguished from conformity viewed as normative or typical behavior, as in Floyd Allport's classical J-curve studies (1).

Although a survey of the literature revealed no direct investigation of problem solving and conformity, indirect but relevant information is found in the fact that similar personality traits correlate with both. Studies by Rokeach (15) and Brown (4) have demonstrated that rigidity in problem solving, measured by the Luchins water-jar test, is related to such variables as ethnocentrism and authoritarianism as measured by the California E and F scales respectively.[2] Conformity, too, has been shown to be significantly related to these same or similar variables in experiments by Crutchfield (7) and Hoffman (9). These studies provide support for the assumption that rigid problem solvers and high conformers have certain personality characteristics in common.

The present study was designed to test three hypotheses. The major hypothesis posits a negative correlation between tendency to conform and achievement in problem solving, when the influence of intelligence is statistically removed. The use of partial correlation permits a demonstration that conformity is associated with a portion of the variability in problem-solving performance that is not accounted for by differences in intelligence as measured by the conventional tests. It is generally accepted that general intelligence is a factor in problem-solving ability. Crutchfield (7) has reported a correlation of −.51 for a verbal test of intelligence and conformity.

The second hypothesis states that the negative partial correlation is higher for problems that require restructuring than for straightforward problems in which restructuring is unnecessary. This elaboration is plausible if one assumes that the water-jar Einstellung task—performance on which has indirectly been related to conforming behavior—has more in common with the restructuring type of problem than with the straightforward type, in that the two former require the modification of an initial approach in order to arrive at an adequate solution. This flexibility in approach requires additionally a critical evaluation of the information presented. It is assumed that these characteristics are also involved in the process of making independent judgments in the conformity situation.

---

[1] This is a condensation of a report which was part of a general program of research on thinking in the Department of Psychology at Stanford. It was carried on under Project NR 150-149 and partially supported by Contract N6onr 25125 between Stanford University and the Office of Naval Research. Permission is granted for reproduction, translation, publication, use, and disposal of this article in whole or in part by or for the United States Government. The original report has been submitted to the committee on Graduate Study of Stanford University as a dissertation in partial fulfillment of the requirements for the Ph.D. in psychology. The writer is indebted to Donald W. Taylor for his support throughout the study. Thanks are due to Kenneth B. Little and Quinn McNemar for their suggestions and to Olga W. McNemar for editing the original manuscript. The writer is especially grateful to Richard S. Crutchfield for making available the research facilities of the Institute of Personality Assessment and Research, University of California, Berkeley.

[2] Since this study was designed and carried out, Levitt (11), in a recent review of the literature, has presented strong evidence that the validity of the water-jar test as a measure of rigidity is highly questionable.

The third hypothesis is that the sex differences found for certain types of problem solving are associated with sex differences in conformity. Several experiments (3, 8, 16) have demonstrated that men are superior to women in particular types of problem solving. Sex differences in conformity (7, 17) and in suggestibility (5) have also been demonstrated.

METHOD

Subjects

The *S*s were 64 men and 77 women students in the introductory psychology course at the University of California, Berkeley. The age range was 16 to 21 years for the women and 18 to 36 years for the men with respective median ages of 18 and 19.

Measures

Problem solving. The problem-solving measure consisted of 20 problems in booklet form. On the basis of an item analysis, the individual problems for this test were selected from those originally used by Sweeney (16) and Hunt and Taylor (10), so as to provide measures of achievement on two types of problems, restructuring and straightforward. Problems of the latter type were workable by direct methods. The former were problems in which for most *S*s the initial set or perception of the problem was inappropriate and a change in set or restructuring was required before solution could be attained. Ten straightforward and ten restructuring problems were arranged alternately and in order of difficulty. A working time of four minutes per problem was allowed with a ten-minute intermission after the first ten problems. Problems were scored either correct or incorrect, with a correct answer receiving one point credit. No partial credits were given.

Conformity.[3] The scale used consisted of 97 items mounted on individual slides. There were 52 critical items which were divisible into the following categories:

24 objective items; 4 each of 6 types
16 attitude items; 4 each of 4 types[4]
12 preference items; 4 each of 3 types.[5]

A response was scored as conforming if it deviated from the objectively correct response in the direction of the supposed responses of the other members of the group. A conforming response received one point and other responses none. The maximum possible score thus was 24, the number of items on the objective conformity scale.

The apparatus for measuring conformity involved five adjacent electrical panels on each of which were five rows of 11 signal lights and a row of 11 switches. The five rows of lights were lettered A, B, C, D, and E. The panels were separated by side wings which prevented each *S* from seeing any panel except the one in his own booth. The five panels were wired separately to the *E*'s control panel, located behind the *S*s, in such a way as to permit the *E* to send identical information to all five panels simultaneously. The *S*s' responses were received on a section of the *E*'s control panel by way of five rows of 11 signal lights, each row corresponding to a booth.

General Intelligence.[6] Part II, Analogies of Form B of the Concept Mastery Test[7] was selected as the measure of general intelligence. There are 75 multiple-choice analogy items; the score is the number right minus one-half the number wrong. The test was administered during one of the regularly scheduled lecture periods.

Procedure

All the students in the class of approximately 650 were informed that they were required to participate as *S*s in two sessions of psychological experiments. For this experiment, *S*s registered for a two-hour first session in groups no larger than 25 persons. The groups were composed without regard to sex, and the *S*s were seated around tables partitioned so that they could not see the papers of other *S*s. At the end of this first meeting, devoted to the problem-solving task, *S*s were asked to sign up for a second session, approximately an hour and a half in length, in groups of five men or five women.

---

[3] Except for slight modification of the content and the order of presentation of the items, the method, materials and apparatus for measuring conformity were those designed and described by Crutchfield (7). They were used for this experiment with his cooperation.

[4] No further description of the attitude items will be presented since scores on them were not included in the final analysis of the data for reasons to be given below. A detailed description of the items, procedure of administration and system of scoring can be found in the original report (14). Copies of this report and also of that by Carey (6) and by Sweeney (16) are available in each of more than twenty major university libraries.

[5] These items are not included in the analysis of this experiment since the "normally" expected responses to them had not been established. The items were used by Dr. Crutchfield who employed the same subjects for another experiment.

[6] There was also available a score for each *S* on Form A of the College Vocabulary Test. This test was developed by H. G. Gough, Institute of Personality Assessment and Research, University of California, Berkeley. The relation of these scores to some of the experimental variables was included in the original report (14) for comparative purposes. However, the results will not be presented here since in the major analysis of the data, only the Analogies test scores were treated as the measure of general intelligence. The correlation between the two tests was .77 for the men and .68 for the women.

[7] The Concept Mastery Test was used with the kind permission of Lewis M. Terman, Stanford University.

At the second session, the *S*s were seated at their respective panels in individual booths. The procedure was explained to them in some detail. About midway of the instructions, the *S*s were shown five cards each lettered with an A, B, C, D, or E. They were told that they were to draw cards to determine the order in which the five *S*s were to judge the stimuli as they were projected onto the screen.

The card-drawing took place immediately preceding the presentation of the first item to be judged. Before the *S*s drew for letter designations, the cards were covertly switched, so that every card carried the letter E and each *S* was thus led to believe he was to respond last in the E position. The responses he saw appearing in the rows on his panel supposedly corresponding to *S*s A, B, C, and D were actually predetermined, simulated responses transmitted by the experimenter from the control panel. The five *S*s thus experienced identical stimulus conditions and gave their responses at the same time. To illustrate with one of the objective items: A slide was projected on the wall in front of the five *S*s, containing a circle and a star and instructions to judge which figure had the greater area. Each *S* saw the other four simulated responses—all indicating the star as larger in area—before giving his own response. These simulated judgments were contrary to each *S*'s perception, since the circle was in fact about one-third greater in area than the star. On all critical slides, pressure to conform was induced by presenting a majority consensus that was inconsistent with the *S*s' perceptions.

At the end of the second session, the experimenter discussed the study with the *S*s, explaining the nature of the deception.

RESULTS AND DISCUSSION

The reliability coefficients, means, and standard deviations for the problem-solving and conformity measures for men and women are presented in Table 1. Considering the nature of the problem-solving test, the reliability coefficients of .70 and .76 for total scores were about as high as could be expected, although higher reliability for the restructuring and straightforward subtests would have been desirable. The reliability values for the objective conformity scores, .87 and .80, were only slightly lower than the .90 reported by Crutchfield (7) for total scores including not only objective, but also attitude and preference items. However, the reliabilities of .29 and .13 obtained for men and women respectively on the attitude conformity items were so low that no further analysis of data based on these items appeared justified. On the basis of unpublished data obtained earlier by Crutchfield, it had been expected that such attitude items would yield a considerably more reliable scale. Just why this expectation was not confirmed is not clear. In any case, all subsequent analyses were limited to conformity as measured by the objective items.

The correlation between the restructuring and the straightforward problems was .59 for men and .58 for women.

Within Groups Analysis

Table 2 presents the correlations among the measures of problem solving, conformity, and general intelligence.

TABLE 1

RELIABILITY OF THE PROBLEM SOLVING AND CONFORMITY MEASURES

(64 men and 77 women)

Measure	Group	*r*	*M*	*SD*
Total problems	M	.70	8.83	3.78
	W	.76	6.00	3.45
Restructuring problems	M	.54	4.33	2.16
	W	.60	2.79	1.95
Straightforward problems	M	.56	4.50	2.08
	W	.52	3.21	1.95
Attitude conformity	M	.29	8.63	3.11
	W	.13	9.75	3.17
Objective conformity	M	.87	8.44	5.09
	W	.80	10.39	4.70

Note.—Reliability was evaluated by the odd-even method for the total problems and the objective conformity scores. The first half versus the second half method was used for the attitude conformity scores. The *r* values were corrected for length by the Brown-Spearman formula. Reliabilities for the restructuring and the straightforward problems were computed by the Kuder-Richardson formula.

TABLE 2

CORRELATIONS BETWEEN MEASURES OF PROBLEM SOLVING, CONFORMITY, AND GENERAL INTELLIGENCE

(64 men and 77 women)

Measure	Objective Conformity		Analogies	
	M	W	M	W
Total problems	−.31**	−.18	.35**	.26**
Restructuring problems	−.38**	−.14	.35**	.22*
Straightforward problems	−.19	−.18	.27*	.25*
Analogies	−.31**	−.08		

* Significant at the .05 level.

** Significant at the .01 level.

Note.—The *p* values are for a one-tailed test since the directions of the correlations were predicted. A one-tailed test was used in all of the subsequent tables, except where otherwise indicated.

For men, conformity was found to be significantly negatively correlated both with total problems and with restructuring problems, but not with straightforward problems. After intelligence was partialled out, the correlations for conformity with restructuring problems (−.28) and with total problems (−.22) were significant at the .01 and the .04 levels. The z' transformation was used to test the significance of these partial correlations. These results support the first hypothesis: there is a negative correlation between tendency to conform and achievement in problem solving when the influence of intelligence is statistically removed. This finding indicates that conformity accounts for variation in problem-solving scores in addition to that which can be attributed to differences in intelligence.

For women, the zero-order correlations between problem solving and conformity were negative; they were, however, not significantly different from zero and hence cannot be regarded as supporting the research hypothesis. On the other hand, these correlations do not provide grounds for rejecting the hypothesis in that they did not differ significantly from those for the men.

The results for the men, for whom the restructuring problems were significantly negatively related to conformity whereas the straightforward problems were not, are compatible with the second hypothesis, that the negative partial correlation is greater for the restructuring problems than for the straightforward problems. However, the difference between the partial correlations was not significant ($p = .09$). Negative results were also obtained for the women.

An inspection of the scores for men revealed that the older men tended to score higher on problem solving and analogies but lower on conformity than the younger men in the sample. For this reason, separate analyses were made for the group of fifty men 21 years of age or younger and for the group of 14 men ranging in age from 22 to 36. These results are shown in Table 3. For the fifty younger men, only the correlation between restructuring problems and conformity was significant. On the whole, the magnitude of the correlations for the younger men differed little from that for the women.

For the 14 men aged 22 and over, there were high negative correlations between problem solving and conformity. After the effect of intelligence was partialled out, the correlation was −.61 ($p = .02$) for total problem solving versus conformity and − .76 ($p = .005$) for restructuring problems versus conformity. With the influence of intelligence and age controlled simultaneously, the respective partial correlations were −.59 ($p = .03$) and −.77 ($p = .005$). The z' technique was used to test the significance of these partial correlations.

The results in Table 3 show also that the difference between the correlations of restructuring and straightforward problems with conformity was in the hypothesized direction. However, the difference between the partial correlations was again not significant. For both the young and the older groups p was larger than .10. Thus, the analysis of the scores for the older men clearly supports the first hypothesis of the study but not the second hypothesis.

Between Groups Analysis

The third hypothesis was that sex differences in certain types of problem solving are associated with sex differences in conformity. It was reasoned that cultural factors that impose differential experiences, training, pressures, and expectations would result in differences in motivations and skills for the two sexes. It was expected that these differences would be manifested in performance on certain problem-solving tasks and under con-

TABLE 3

CORRELATIONS BETWEEN MEASURES OF PROBLEM SOLVING, CONFORMITY, AND GENERAL INTELLIGENCE

(50 young men [Y] and 14 older men [O])

Measure	Objective Conformity		Analogies	
	Y	O	Y	O
Total problems	−.20	−.70**	.25*	.65**
Restructuring problems	−.25*	−.79**	.21	.70**
Straightforward problems	−.11	−.59**	.25	.38
Analogies	−.22	−.44		

* Significant at the .05 level.

** Significant at the .01 level.

Note.—The z transformation was used to test the significance of the correlations for the group of 14 older men.

formity conditions similar to those of this experiment.

For this hypothesis to be valid first requires that men be superior to women in certain types of problem solving and at the same time show less tendency to conform. Table 4 shows that these prerequisite conditions were clearly fulfilled. The sex differences for all three problem-solving scores were highly significant. However, the sex difference on restructuring problems was not significantly greater than that on the straightforward problems. The sex difference on the conformity measure was significant as expected. Surprisingly, a significant sex difference was also obtained on general intelligence. Because of this unexpected difference, it was necessary to adjust for both intelligence and conformity in carrying out the analysis of covariance designed to test the hypothesis. The results of the analyses, carried out separately for the restructuring and the straightforward problems, are presented in Table 5. Although there was some reduction in the differences, both *F* ratios were still significant at the .01 level, indicating that not all of the sex differences in problem solving was accounted for by sex differences in conformity and intelligence as measured.

A similar analysis of covariance was not carried out for the sex difference on total problems. Since significant differences remained after adjustments on the restructuring and the straightforward problems separately, it was clear that the difference would be significant for total problems should adjustments be made for conformity and intelligence.

The question may be raised as to the relevance of the results of the separate covariance analyses of the restructuring and straightforward problems to the earlier findings of Sweeney (16). In a series of experiments Sweeney found that for groups matched with respect to intelligence, special abilities, or specific knowledge, men were superior to women on restructuring types of problems but not on straightforward problems. In the present experiment men proved to be superior to women in both types of problems after the covariance adjustments were made for conformity and intelligence. Though these results do not confirm Sweeney's findings, it cannot be said that they are contradictory since the groups in the present study were not equated on relevant special abilities or on related specific knowledge.

TABLE 4

SEX DIFFERENCES

(64 men and 77 women)

Measure	Group	*M*	*SD*	*D*	*p*
Total problems	M	8.83	3.78		
	W	6.00	3.45	2.83	<.001
Restructuring problems	M	4.33	2.16		
	W	2.79	1.95	1.54*	<.001
Straightforward problems	M	4.50	2.08		
	W	3.21	1.95	1.29*	<.001
Conformity	M	8.44	5.09		
	W	10.39	4.70	1.95	.01
Analogies	M	37.19	12.17		
	W	29.58	10.68	7.61	<.001[a]

* The difference between mean differences in restructuring problems and straightforward problems is .25; *p* is not significant.

[a] The *p* value is for a two-tailed test since the difference was not predicted.

TABLE 5

ANALYSES OF COVARIANCE: SEX DIFFERENCES ON RESTRUCTURING PROBLEMS AND ON STRAIGHTFORWARD PROBLEMS, SEPARATELY, WITH ADJUSTMENTS FOR CONFORMITY AND ANALOGIES SCORES

Test	Source	Before Adjustment		After Adjustment			
		df	Mean Square	*df*	Mean Square	*F*	*p*
Restructuring problems	Total	140	4.79	138	4.00		
	Within	139	4.24	137	3.79		
	Between	1	82.00	1	32.44	8.56	.01
Straightforward problems	Total	140	4.41	138	3.99		
	Within	139	4.09	137	3.79		
	Between	1	58.00	1	30.71	8.10	.01

A further analysis of sex differences on problem solving was made to determine whether or not sex differences would remain if the older men were excluded. The differences between the means for 77 women and 50 younger men were computed for the problem solving, conformity, and analogies scores. The differences on all three measures of problem solving remained significant at the .001 level. The scores of the older men were thus not entirely responsible for the obtained sex differences. For conformity, the difference was reduced in level of significance from .009 to .07. The difference on analogies was still significant at the .003 level. An analysis of covariance with adjustments for conformity and intelligence was not carried out for this comparison as it was for the difference in problem solving for the total men and the women. Since the difference between the means for the younger men and the women in conformity was very small and since the correlations between conformity, problem solving, and intelligence scores for the younger men were less than the corresponding correlations for the total men, covariance adjustments would not be expected to reduce the sex difference in problem solving to insignificance.

SUMMARY

This study was undertaken to investigate the possibility that conformity, a nonintellectual variable, contributes to the variability in achievement in problem solving in addition to that variation accounted for by intelligence as measured by the conventional tests. Three hypotheses were tested:

1. There is a negative correlation between tendency to conform and achievement in problem solving when the influence of intelligence is statistically removed.

2. The negative partial correlation is higher for problems that require restructuring than for straightforward problems.

3. The sex differences in certain types of problem solving are associated with sex differences in conformity.

The *S*s, 77 women and 64 men, were given tests of problem solving, conformity, and intelligence. The problem-solving test consisted of 10 restructuring and 10 straightforward problems. Part II, analogies, of Form B of the Concept Mastery Test was used as the measure of intelligence. Two measures of conformity were used, one involving judgments concerning objective stimuli and the other judgments concerning attitudes. Since the reliability of the latter proved to be extremely low, all subsequent analyses were limited to conformity as measured by the objective items.

The results supported the first hypothesis but failed to confirm the second and third hypotheses.

REFERENCES

1. ALLPORT, F. H. The J-curve hypothesis of conforming behavior. *J. soc. Psychol.*, 1934, **5**, 141–183.
2. ASCH, S. E. Studies of independence and conformity: 1. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, No. 9 (Whole No. 416).
3. BILLINGS, M. L. Problem solving in different fields of endeavor. *Amer. J. Psychol.*, 1934, **46**, 259–272.
4. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
5. BROWN, W. Individual and sex differences in suggestibility. *Univer. California Publ. Psychol.*, 1916, **2**, 291–430.
6. CAREY, G. L. *Reduction of sex differences in problem solving by improvement of attitude through group discussion.* Stanford: Dept. Psychol., Stanford Univer., 1955 (Tech. Rep. No. 9, Contract N6onr 25125).
7. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
8. GUETZKOW, H. An analysis of the operation of set in problem solving behavior. *J. gen. Psychol.* 1951, **45**, 219–244.
9. HOFFMAN, M. L. Some psychodynamic factors in compulsive conformity. *J. abnorm. soc. Psychol.*, 1953, **48**, 383–393.
10. HUNT, E. B., & TAYLOR, D. W. Personal communication. January, 1955.
11. LEVITT, E. E. The water-jar Einstellung test as a measure of rigidity. *Psychol. Bull.*, 1956, **53**, 347–370.
12. LUCHINS, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).
13. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
14. NAKAMURA, C. Y. *The relation between conformity and problem solving.* Stanford: Dept. Psychol., Stanford Univer., 1955 (Tech. Rept. No. 11, Contract N6onr 25125).
15. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
16. SWEENEY, E. J. *Sex differences in problem solving.* Stanford: Dept. Psychol., Stanford Univer., 1953 (Tech. Rep. No. 1, Contract N6onr 25125).
17. WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

Received May 13, 1957.

A RESTATEMENT AND TEST OF SCHLOSBERG'S THEORY OF EMOTION WITH TWO KINDS OF SUBJECTS FROM GREECE[1]

HARRY C. TRIANDIS
Cornell University
AND WILLIAM W. LAMBERT
Center for Advanced Study in the Behavioral Sciences

MANY propositions in the literature of psychology are stated in such a general form that replication with human subjects from very diverse cultural backgrounds is directly relevant. In the present state of our understanding, such replication is probably required whenever a proposition reflects experimental operations dependent upon the more subtle uses of perceptual or language responses. If the proposition is upheld by the test, then the possibility that the relation depends on rare cultural or social factors is rendered less likely.

This study represents an attempt to test Schlosberg's dimensional theory of emotion (15) and to compare the relationships found using American college *S*s with those found with Greek college students from Athens and *S*s from a small Greek village. Besides the test of the main hypotheses relating to similarities in the data, certain exploratory hypotheses concerning differences are also discussed.

Schlosberg (14) has attempted to reduce the dimensions required to order judgments of pictures that presumably reflect "emotional states" in the person photographed. Woodworth (17) had *S*s place pictures into categories named: 1. love, happiness, mirth; 2. surprise; 3. fear, suffering; 4. anger, determination; 5. disgust; 6. contempt. Schlosberg has shown that it is reasonable to view this continuum as a circle rather than a straight line. Further, he has shown that there is a high correlation between the way *S*s place the pictures into Woodworth's six categories and the way they place the same pictures according to the two bipolar nine-step dimensions of: 1. pleasant-unpleasant; 2. attention-rejection. These dimensions are fewer in number than those of Woodworth and they possibly involve somewhat less complex judgments. However, much of their face validity and their relation to the extensive everyday language realm of "emotional states" depends largely on the even more subtle and complex discriminal and symbolic processes involved in the use of the Woodworth categories. More recently Schlosberg (15) has extended his theory to include a further dimension, tension-sleep. The appropriate model thus becomes an inverted ellipsoidal cone, with "sleep" anchoring the base point of the cone.

Although this latest version of the theory is still primarily concerned with the *perception* of emotional expression, the physiological indices used by Schlosberg for the measurement of the tension-sleep dimension suggest that the theory is a first step towards a general theory of emotion. One can think here of Hebb's (2), and of Lindsley's (4) discussion of the differential firings in the reticular formation as the physiological changes underlying the *sleep-tension* dimension, of Olds's (9) firing of the "reward system in the brain" (the septal area facilitation and inhibition) which suggests the basis for the *pleasant-unpleasant* dimension, and of the action of the sympathetic and parasympathetic nervous systems, with their dilation and constriction of the pupil of the eye and the bronchi etc., which suggest a basis for the *attention-rejection* dimension. In short, although the present data involves judgments of facial expressions, it may be possible to generalize the *dimensions* abstracted from these data to subsume a good portion of the variance of the actual emotions people feel. Certainly when one judges the emotion shown in a face, the particular judgment may not correspond either to one's own emotion at the time or to that of the person who is moving his face. The dimensions of the judgment, however, may turn out to be adequate to describe both one's own and the other's emotional states. Proof of such an hypothesis, however, will depend on both the development of more data on judgments as well as more independent physiological measurements, and, most important of all, development of experiments relating both of these.

---

[1] This study was assisted in part by a grant from the U. S. Public Health Service to the second author and by the general research funds of the Department of Psychology, Cornell University. Writing and analysis was in part facilitated by the Center for Advanced Study in the Behavioral Sciences. The second author is on leave from Cornell University. The authors are appreciative of suggestions given by A. T. Ryan and J. E. Hochberg.

Osgood (11) has recently reported the results of a factor analysis of *S*s' responses to forty posed facial "expressions," using the semantic differential as a measurement device. He suggests that the factors he named after applying Thurstone procedures to the picture judgments—1. pleasantness; 2. control; 3. expressiveness;—"look like" the major dimensions of his "semantic space"—1. evaluation; 2. potency; 3. activity. This appears to be a reasonable suggestion, and it only remains for us to suggest that Schlosberg's three dimensions appear to be either a special case or a biologically more generic case of the dimensions later named by Osgood.[2]

Keeping in mind the problems of factor naming, or even of dimension naming, one may suggest that Schlosberg's "pleasant-unpleasant" sounds like Osgood's "pleasantness" and "evaluation," and "tension-sleep" sounds like "activity" and "expressiveness," whereas "attention-rejection" sounds like "potency" (in the sense of potency of the cause of emotion) and more clearly, "control."

If such convergence turns out to be the case it would be interesting. A rather nice question would then arise as to which set of operations might be most profitably viewed as more generic. The facial expression operations have one advantage over the more complicated "semantic" ones in that they keep open the possibilities of a direct psychophysical approach to the further clarification of the matter. Although symbol capacity and usage is certainly of vital importance in human behavior, the basic and most general dimensions of such capacity and usage may be presaged in some psychophysical predispositions common with some other animals and with all peoples. Teasing out the parameters involved in the stimuli of "emotional expressions" may be the most direct attack on the bases of the more elaborated and secondary semantic mediating processes.

Since there is no basis a priori for assuming that cues for discriminating "facial expressions" are more universal or generic than the apparently more complex semantic dimensions, both formulations require tests in very different cultural milieu before any general formulation of either theory be taken very seriously. Kumata (3) and Triandis and Osgood (16) have attempted cross-cultural tests of the semantic operations and their attendant hypotheses. The present paper reports such an attempt for the facial "expressions" operations. As a result of such attempts we may be able to take seriously the partial but general pattern that seems to be emerging from all of these studies.

This possibility is all the more interesting when the similar convergence in the independent work of Morris (7) on dimensions of "values" and their setting in a theory of behavior (8) is considered. Morris and Jones (6) have factor analyzed the relations between the ratings of thirteen "ways of life." They report five Thurstone factors, three of which can be viewed as relevant here. Their "receptivity and sympathetic concern" factor "looks like" a meeting of the "sleep" or "passive" end of the "sleep-tension" dimension and of the "acceptance" end of the "acceptance-rejection" dimension. Their "enjoyment and progress through action" factor could be a meeting of the "active" and "approach" ends of the Schlosberg dimensions, and finally, their "withdrawal and self-sufficiency" factor could be a meeting of the "withdrawal" and "rejection" ends of our continua.[3] This suggestion, of course, does no more than outline the *possibility* of a convergence of the Schlosberg dimensions and of the Osgood "meaning" dimensions with basic dimensions of "valuing" behavior.

METHOD

Subjects

The American data have been kindly provided by T. Engen, N. Levy, and H. Schlosberg, based on ninety Brown University students. Two groups were tested in Greece during the summer of 1956. Fifteen *S*s were studied in the village of Sfakera on the northeast coast of the island of Corfu, and fifteen college students from Athens. In both groups there were almost an equal number of *S*s of each sex. Other characteristics of the two groups are given in Table 1, which makes clear that there are large differences between the groups in

---

[2] The editor of this Journal has suggested that our formulation is quite parallel to Wundt's tridimensional theory of feeling.

[3] These suggestions were made by Fred Attneave in personal communication.

TABLE 1

CHARACTERISTICS OF SAMPLES

Statistic	Age		Education (years)		Movie Attendance (lifetime)	
	V	A	V	A	V	A
Mean	18.6	21.1	7.1	14	19	1000
Median	19	18	6.0	13	12	900
Range	15–24	17–33	6–11	11–18	0–50	100–3000

Note.—V stands for village and A for Athens.

educational level and in their attendance at movies—largely the same movies American students would attend. The data on movie attendance in the table represent estimates by the *S*s themselves.

Procedure

The first problem was to ensure equivalence of terms in Greek and English. This was approximated by having a group of bilinguals translate the terms to be used from Woodworth's and Schlosberg's English into Greek, and another group of bilinguals translate the Greek back into English. Only terms which consistently met a criterion of complete agreement both ways were accepted for the Greek version.

In the Schlosberg procedure, anchoring pictures were presented while instructions were given, and then taken away before an *S* began rating. Pictures numbered 8 and 32 of the Ruckmick (13) series were used for the pleasant-unpleasant dimension, Numbers 7 and 13 for the attention-rejection dimension and 26 and 4 for the tension-sleep dimension.

The pictures judged by the *S*s were the new set of pictures of Marjorie Lightfoot, kindly provided by Schlosberg, which are reproduced in Morgan's workbook (5, p. 32–33).

The nine-point scales and the Woodworth six categories were drawn on cardboard for presentation to the *S*s.

The villagers were instructed as follows:

> This is an important study which I am making, but as far as you are concerned it is just a game. All you have to do is to play the game *correctly*. The rules of the game are as follows: you are to place each of these forty-eight photographs in one of these places (show on cardboard). First you will place those that represent a pleasant emotion on *this* part of the cardboard (indicate). Then you will place those that represent an unpleasant emotion on *that* part of the board (indicate). Then you will spread them out, so that the more pleasant ones are there, and they gradually get less and less pleasant, up to here (indicate) and more and more unpleasant from here (indicate) to there.

Similar instructions were given for the other dimensions. In order to make sure that the idea "gradualness" was communicated a pencil was tapped on a table at various intensities. In all cases in the village, the *S* was not out of sight of his peer group, but was sufficiently far from his friends so that they would not influence his ratings. The villagers were very eager to "play the game" and they all said that they "enjoyed playing it." Token gifts were given to all participants and refreshments were served in all instances. This kind of non-business-like atmosphere is thought to be essential for psychological research in Greek villages.

In the case of the city group the instructions were the same, but the task was presented as an experiment, and a test of their ability to recognize emotions.

RESULTS AND DISCUSSION

Similarity Between Ratings of Greek and American Subjects

The first hypothesis was that Greek *S*s rate emotional "expression" in much the same way as American *S*s. A test of this hypothesis requires that there be a significant and fairly high correlation between the way the three groups place the pictures on the three Schlosberg dimensions. Such a test is shown in Table 2. Rank-order correlation coefficients were computed on the ranks of the forty-eight pictures. The lowest correlation coefficient in this table is .66 and is significant at the .0001 level. Hence, there is little doubt that Greek *S*s, even when they come from very different populations, rate emotional expressions in the same way as American college *S*s.

A corollary of this hypothesis is that if the P-U, A-R, and S-T are assumed to be orthogonal, and the position of the photographs plotted accordingly, the *emotion solid* obtained will be similar for the Brown University and Greek data. This hypothesis can be verified by visual inspection of Fig. 1. These solids are plotted by the method of successive intervals, a scaling procedure derived from Thurstone's law of comparative judgment (see Edwards [1, pp. 120–135]).

Briefly, we obtain all available pictures of a certain level of tension, say between 7 and 8 (on a nine-point scale), and note how they have been rated on the A-R, and P-U scales. We then treat each picture just as Edwards

TABLE 2

RANK-ORDER CORRELATION COEFFICIENTS

Groups	Dimension		
	P-U	A-R	S-T
Brown vs. Athens	.91	.82	.76
Brown vs. Village	.82	.88	.67
Athens vs. Village	.86	.83	.66

Note.—Based on ranks of forty-eight pictures showing relations among three groups studied.

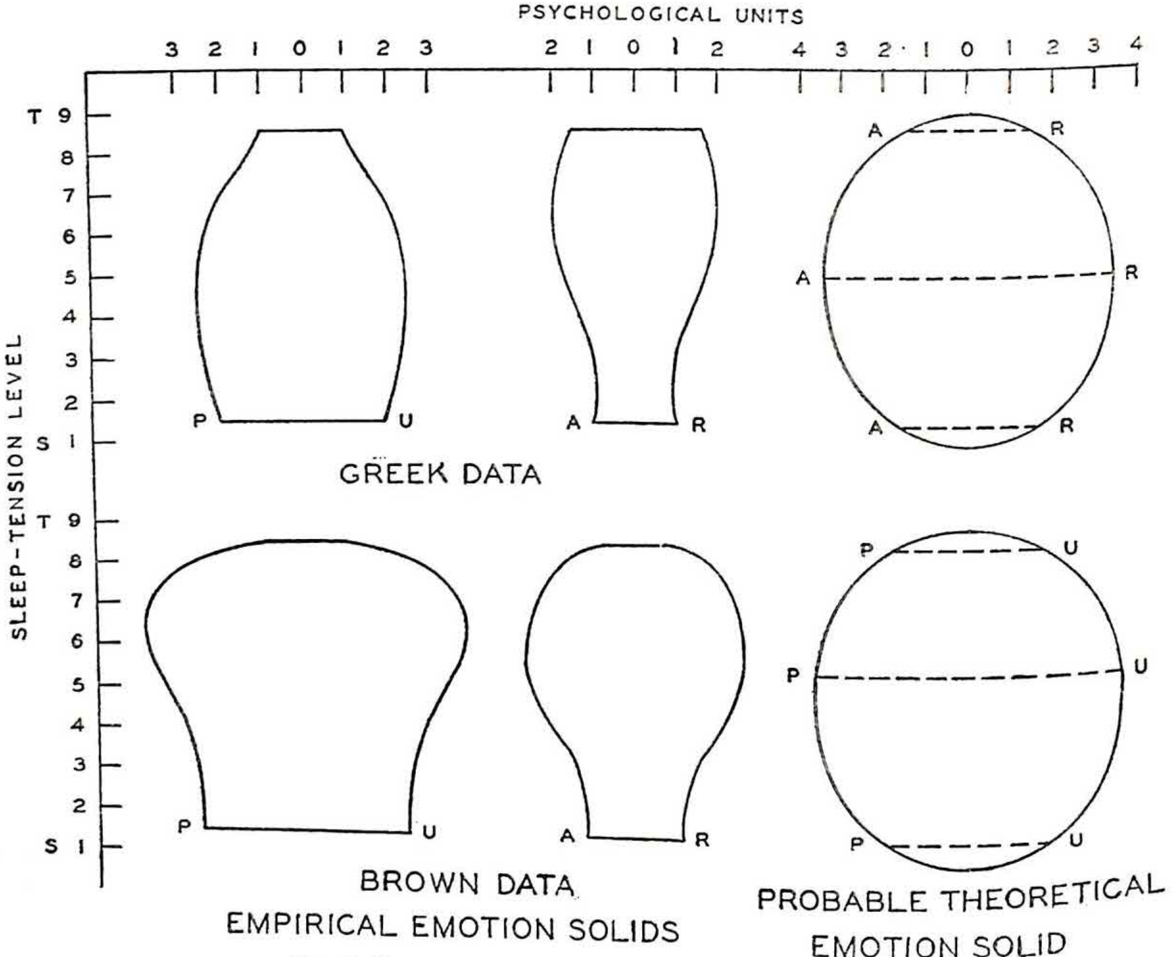


FIG. 1. EMPIRICAL AND THEORETICAL EMOTION SOLIDS

treats a *statement* for attitude scale construction: the frequency of judgments of this picture at a certain point of a psychological continuum is noted, this is converted to a cumulative distribution, and to a cumulative probability score. The latter is transformed to a normal deviate (z score) by referring to the normal probability distribution. The difference between adjacent intervals can thus be obtained. Repetition of this procedure for several pictures and averaging gives an estimate of the interval width for each one of the scale graduations of the P-U and A-R scale. If we add all the intervals, we get the length of the P-U and A-R axes. Since such axis-lengths can be computed at each level of tension we obtain several axis-lengths and use them to plot our solid.

We must not place too much confidence on the actual distances on these solids, since they are based on *only one* picture at the tension level between 1 and 2 and only 4 or 5 pictures for tension levels 3 to 4, 5 to 6, 7 to 8, and 8 to 9. These are small samples indeed of what are probably rather large populations. Furthermore, the American solid is based on more than 90 *S*s, the Greek on only 30 *S*s. Differences in the general width of the irregular ellipsoidal cylinders obtained may be due to better discriminations made by the greater number of *S*s for the American data. Nevertheless, we can cautiously conclude that the emotion solid is probably closer to an irregular ellipsoidal cylinder rather than an inverted ellipsoidal cone as was postulated by Schlosberg.[4] Schlosberg's guess on the shape of the solid was based on theoretical rather than empirical considerations. Our solids are completely empirical.

---

[4] The assertion that fewer discriminations can be made at the ends of the tension continuum than in the middle depends upon acceptance of the following assumptions: (*a*) a normal distribution of judgments for each picture (this is clearly not always true); (*b*) perfect test-retest reliability; (*c*) rating one picture does not influence the rating of another; (*d*) the distance between the poles remains constant throughout the series of judgments; (*e*) the poles mean the same thing to all *S*s.

If a more smooth looking solid is desired, our best guess is to assume that the regular ellipsoid of Fig. 1 will be a reasonable approximation. Much more work is needed before the exact shape of the emotion solid can be determined. A serious attempt will have to be made to construct pictures which push at the outside contours of the solid, and more work must be done on scaling the sleep-tension levels themselves. Though the present method is laborious, it yields results that seem reasonably consistent even with very rough data.

Reliability of Judgment: Schlosberg Dimensions vs. Woodworth Categories

The second hypothesis was that Greek *S*s will place pictures on the Schlosberg dimensions (pleasantness-unpleasantness, attention-rejection, and tension-sleep) more reliably than when they use the Woodworth categories. This hypothesis can only be given a rough test because of the difference in procedure in presenting the scales in the Schlosberg method (two polarities at a time) and that in the use of the Woodworth categories (six categories at once). Results are fairly clear from a careful "visual" check of the data. Two photographs in the case of the village group and one photograph in the Athens group were rated in what seems to amount to a random fashion on at least one dimension, when the Schlosberg procedure is used. In the use of the Woodworth categories, however, 24 of the pictures in the village and 11 in Athens, were allocated in what amounts to a random fashion. These findings suggest that the Schlosberg dimensions permit more reliable ratings of the photographs than do the Woodworth categories.

Predictability of Woodworth Categories from Schlosberg Dimensions

The third hypothesis was that the position in the Woodworth categories can be predicted from its rating on the pleasant-unpleasant and attention-rejection dimensions. We expect this prediction to be better in the case of the Athens group than in that of the village group since the Athens group is closer to the cultural inductive base—the United States—largely because of repeated attendance at American movies. To test this hypothesis the positions of the pictures as determined by the P-U and A-R dimensions were computed and plotted. The surface of the graph was divided into six equal sections, following Schlosberg's procedure. To permit a test of significance, a table such as sketched in Table 3 was prepared. The circularity of the Woodworth categories was assumed, since Schlosberg has shown that this is a reasonable assumption (14). According to Schlosberg, there should be a relationship between the placement in the six sections of the graph and the placement on the six qualitative categories of Woodworth. The "predicted category" is, therefore, the section of the graph in which a picture was placed, and the "mean actual category" is the mean placement in the Woodworth categories. If the actual category differed from that predicted by one or more categories, an error was counted. If it differed by 0.9 or less of a category, no error was counted. By this criterion, if the predictions were done by chance alone, two-thirds of the photographs would produce errors. Actually, fewer errors were produced. A chi-square test of goodness of fit between the predicted and the actual number of errors showed fewer errors with $p < .001$ for the Athens group, $p < .05$ for the village group. We conclude that the Greeks followed Schlosberg's predictions in showing a relationship between their placement of a picture on the P-U and A-R dimensions and in the Woodworth categories. These results provide some support for the third hypothesis.

TABLE 3
ERROR IN PREDICTION OF WOODWORTH CATEGORY PLACEMENT

Picture	Predicted Category	Mean Actual Category	Error
1	1	1.3	0
2	6	4.9	1
3	4	3.1	0
etc.	etc.	etc.	etc.
48	5	6	1

Note.—This is only a partial table. It is used in order to arrive at criterion of error of prediction for placement of Woodworth categories from P-U and A-R placement.

Relative Convergence of Groups with One Another

The fourth hypothesis dealt more specifically with the influence of stereotypy. Since the Athens and Brown *S*s see many of the same movies they ought to converge more with one another in their judgments and ratings of facial expressions than with the village group. A chi-square test was used again. For each picture and each dimension the differences in the mean ratings of the three groups were computed. If the Brown mean for a picture on a given Schlosberg dimension was 2.2, the Athens mean 1.9 and the village mean 2.9, the village *S*s were classified as further from the other two since $2.9 - 2.2 > 2.2 - 1.9$, and the instance scored as one case of convergence between Brown and Athens. If the prediction of Hypothesis 4 is incorrect, there should be as many cases of Brown and Athens converging (16 cases) as there are cases of convergence between the other two possible pairings. Actually, this was not the case (see Table 4). The chi square for each dimension is significant at the .05 level. The overall chi square for the table is 11.4 or $p < .005$.

Thus we conclude that there is some factor bringing the Athens and Brown group closer together. A direct test of the relationship between movie attendance and closeness of judgments to the Brown norms was performed and was in the expected direction but did not quite reach significance.

Effects of Cultural Differences

The fifth hypothesis reflected a general expectation of some consistent and significant differences in judgments of the pictures associated with differences in customs. While the present data cannot exclude the possibility of chance association, such differences suggest interesting hypotheses that could be tested using these procedures in conjunction with sample surveys or anthropological methods.

Role of verbal aggression. It is the custom in the Greek village studied to engage in verbal aggression as a means of entertainment—a kind of aggressive play that may function to express anger but also provide practice in control of aggression. Loud and angry debates are a favorite pastime in Sfakera. It is interesting therefore to note that there are significant differences in judging the picture described by Schlosberg as "Intense anger in argument" (No. 22). Compared to either the Athens or Brown groups, the villagers see the picture as more pleasant and less tense (at the .05 and .01 levels respectively). The pictures included six other attempts to portray anger, and in all but one of these the villagers tended to judge them as more pleasant than did the Brown group. The only reversal with the Brown group was with "mild anger" states that might not reflect the dramatic quality enjoyed in their anger by these people. The differences in tension are less marked, but in the three cases where there are differences they reflect the same direction as found in picture No. 22. Differences between the Athens group and the villagers are not maintained in the other pictures, suggesting that a cultural factor may be operating. The hypothesis emerges, then, that to the degree that anger is displayed under pleasurable circumstances pictures of that state are viewed as more pleasurable and less tense. Such a hypothesis seems applicable to individual differences as well as to group or cultural differences.

TABLE 4

RELATIVE CONVERGENCE OF GROUPS TO ONE ANOTHER

Groups	P-U	A-R	S-T	Total
Brown and Athens converging	24	20	23	67
Brown and Village converging	11	12	14	37
Athens and Village converging	13	16	11	40
Totals	48	48	48	144

Role of bodily exertion. It is reasonable to expect differences in the judgment of pictures depicting physical strain between rural people who do physical toil and college students whose main contact with real strain may be limited to sports and games. Such differences are reflected in judgments of picture No. 30, which the villagers judge as less pleasant and yet less tense than do either of the college groups. There appears to be some cultural factor in addition since the same difference holds significantly for the Athens group compared to the Brown group. Certainly the Greeks generally depend much less on mechanical equipment for getting work done. Adaptation to physical strain may make its exemplification seem less tense, but the ubiquity of hard physical toil and the attendant fatigue should lead to association with pain and unpleasantness. The present data therefore, may provide a rough index of the valuation placed on physical toil and the effort expended, and we hypothesize that the valuation varies negatively with frequency of exhausting labor and positively with frequency of games of physical skill, whereas adaptation to such strain is reflected in the effort or tension dimension. Here again individual differences as well as cultural differences should be reflected.

Relations among dimensions. One additional trend that runs throughout the data on the attention-rejection and tension-sleep dimensions is worthy of mention even though its interpretation is more difficult. On the average, the Greeks (whether urban or village) give ratings more toward "attention" and "tension" for *unpleasant* pictures while American college students give ratings more toward "attention" and "tension" for *pleasant* pictures (see Tables 5 and 6). These findings are significant at the .001 level by chi-square test.

In explanation of this finding we might invoke and apply to emotional expression a stereotype according to which Americans are seen as "activist" and "instrumentally oriented." According to the diagram used by Schlosberg (14, p. 127) for representing the relation between his dimensions and the Woodworth categories, the Americans, by giving higher "attention" ratings than the Greeks to pleasant pictures, shift them from the more purely emotional categories of "love, mirth, happiness" toward the more instrumental "surprise." Conversely, by giving relatively less "attention" ranging to unpleasant pictures, the Americans shift the latter from the more purely emotional categories of "fear, suffering" toward the instrumental categories of "anger and determination." The Greeks may be more ready to accept an emotion without having any idea of what to do about it. There is a suggestion that Americans tend to give higher "tension" scores on the average than do the Greeks in the present study.

TABLE 5

Number of Pictures Relatively High on "Attention"

	Americans	Greeks	Total
Pleasant Pictures	12	3	15
Unpleasant Pictures	6	21	27
Totals	18	24	42

TABLE 6

Number of Pictures Relatively High on "Tension"

	Americans	Greeks	Total
Pleasant Pictures	17	0	17
Unpleasant Pictures	9	15	26
Totals	26	15	41

One rather jarring note emerges from consideration of Picture No. 51. Seven out of fifteen villagers perceived it as "extremely pleasurable" while another six scored it as "extremely unpleasant." Both Athens and Brown groups saw it overwhelmingly as unpleasant (scores 1 to 3). This finding actually follows from the shape of the emotion solid. When an axis is shorter at one level of tension than at another discrimination at that level should be more difficult and confusion more likely.[5] Such confusion was typical in the case of extreme judgments of pictures that are judged to be toward either high tension or sleep.

SUMMARY

In the present paper we have shown that the conclusions presented by Schlosberg (1952, 1954) for Brown University subjects are valid in Greece as well. Schlosberg's method can now be used with non-Western groups, to establish the differences between the non-Western and Western modes of perceiving emotional expressions.[6]

Secondly, we have shown that the emotion solid can be plotted from data collected by making use of Schlosberg's procedure, and the shape of the solid is *not* the same as that presented in Schlosberg's 1954 paper.

Thirdly, we have shown that the Schlosberg technique can be used to study certain differences between Western cultural groups.

---

[5] This interpretation may be too simple. There is a highly significant difference in the greater number of bimodal distributions of judgments at the two extreme ends of the sleep-tension continuum as compared to the middle ranges. This fact, together with the statement of some *S*s that extremely pleasant and extremely unpleasant expressions look the same leads to the possibility that stimuli are more ambiguous at extreme tension and low tension levels and *also* that *S*s tend *not* to use the middle categories of the scales at these levels. Serious validation of these findings might suggest a more complicated solid than the one hypothesized in Fig. 1.

[6] Such a study is being planned for a village in Taiwan, to be conducted by Arthur and Margery Wolf.

REFERENCES

1. EDWARDS, A. L. *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts, 1957.
2. HEBB, D. O. Drives and C.N.S. *Psychol. Rev.*, 1955, **62**, 243–254.
3. KUMATA, H., & SCHRAMM, W. A pilot study of cross-cultural meaning, *Public Opin. Quart.*, 1956, **20**, 229–237.
4. LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: John Wiley, 1951. Pp. 473–516.
5. MORGAN, C. T. *Student's workbook to accompany introduction to psychology.* New York. McGraw-Hill, 1956.
6. MORRIS, C., & JONES, L. V. Value scales and dimensions, *J. abnorm. soc. Psychol.*, 1955, **51**, 523–535.
7. MORRIS C. *Varieties of human value.* Chicago: Univer. of Chicago Press, 1956.
8. MORRIS C. Signs, language and behavior. Paper read at Center of Advanced Study in the Behavior Sciences, January, 1957.
9. OLDS, J. Physiological mechanisms of reward. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. of Nebraska Press, 1955.
10. OSGOOD, C. Behavior theory and the social sciences, *Behav. Sci.*, 1956, **1**, 167–185.
11. OSGOOD, C. E. Fidelity and reliability. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill: Free Press, 1955.
12. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana, Ill.: Univer. of Illinois Press, 1957.
13. RUCKMICK, C. A. A preliminary study of the emotions. *Psychol. Monogr.*, 1921, No. 136.
14. SCHLOSBERG, H. The description of facial expressions in terms of two dimensions. *J. exp. Psychol.*, 1944, **44**, 229–237.
15. SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.
16. TRIANDIS, H. C., & OSGOOD, C. E. A comparative factorial analysis of semantic structures in monolingual Greek and American college students. *J. abnorm. soc. Psychol.*, in press.
17. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.
18. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

Received May 20, 1957.

PSYCHOLOGICAL TEST AND ORGANISMIC CORRELATES OF INTERVIEW INTERACTION PATTERNS[1, 2]

RUTH G. MATARAZZO, JOSEPH D. MATARAZZO, GEORGE SASLOW, AND JEANNE S. PHILLIPS

University of Oregon Medical School

In several previous papers, we have shown that certain behavioral variables which characterize an individual's interview interaction pattern during a partially standardized interview can be objectively and reliably assessed by the Interaction Chronograph method (7, 12, 13). This method of assessing personality through interview interaction patterns is based on the assumption that *how* a person interacts with an interviewer in regard to certain temporal interaction measures specified, for example, in terms of frequencies and durations of action, silence, and other variables, may be both characteristic and meaningful. That is, how often a person speaks, how long are his utterances, how long is his reaction time before responding to the interviewer, how often he takes the initiative when the interviewer is silent, etc., may reflect significant and predictive individual differences.

As described by Chapple (3) and by the writers (8, 12), the interviewer's behavior (but not the content of his communications) is standardized during the total interview, which consists of five subperiods. During Subperiods 1, 3, and 5, the interviewer's behavior consists of nondirective, free, give-and-take interviewing. Periods 2 and 4 are stress phases of the interview, which require the interviewer to remain silent following the patient's last utterance (Period 2) and to interrupt the patient while he is talking (Period 4) as a "dominance-submission" situation. The variables measured during the interview are defined as follows: (*a*) *Pt.'s Units*, the number of times the patient acted; (*b*) *Pt.'s Action*, the average duration of the patient's actions; (*c*) *Pt.'s Silence*, the average duration of the patient's silences; (*d*) *Pt.'s Tempo*, the average duration of each action plus its following inaction, as a single measure; (*e*) *Pt.'s Activity*, the average duration of each action minus its following inaction, as a single measure; (*f*) *Pt.'s Adjustment*, the durations of the patient's interruptions minus the durations of his failures to respond, divided by Pt.'s Units; (*g*) *Dr.'s Adjustment*, the durations of the interviewer's interruptions minus the durations of his failures to respond, divided by Pt.'s Units; (*h*) *Initiative*, the per cent of times, out of the available number of opportunities (usually 12) in Period 2, in which the patient acted again (within a 15-second limit) following his own last action; (*i*) *Dominance*, the number of times in Period 4 that the patient "talked down" the doctor minus the number of times the doctor "talked down" the patient, divided by the number of Pt.'s Units in that period; (*j*) *Pt.'s Synchronization*, the number of times the patient interrupted or failed to respond to the doctor, divided by the number of Pt.'s Units; (*k*) *Quickness*, the average length of time in Period 2 that the patient waited before taking the initiative following his own last action; (*l*) *Number of Interruptions*, the number of times the patient interrupted the doctor; (*m*) *Length of Interview*, the duration of the interview in minutes.

The above-defined variables were found to be characteristic of a given patient, in that they showed high reliability (test-retest stability) when he was interviewed by two different doctors a few minutes apart on the same day (7, 12) and when he was interviewed twice by one doctor over a seven-day interval (13). In studying the reliability of the input of the observer who records the raw data from which the variables are printed by the Interaction Chronograph, it was found that the observer's task is a highly reliable, primarily mechanical one, and that neither his memory nor a constant response-set is a probable explanation of the high reliabilities of the interviewee interaction variables (10). Although the measurements of a given patient's behavior made on one occasion were highly reproducible when he was re-interviewed, the values of each interaction variable differed widely from one patient to another. In other words, there were

---

[1] This investigation was supported by a research grant (M-1107) from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

[2] The data were collected while the authors were at the Massachusetts General Hospital and Harvard Medical School.

large interindividual differences but small intra-individual variability in interviewee interaction patterns. The purpose of the present study was to explore what meaning could be attributed to these individual differences in terms of concurrent validity, i.e., the correlation between interview response behavior and other response behavior. This study represents a first approach, designed to provide leads for further investigation of the usefulness of the Interaction Chronograph in making meaningful differentiations among people or in predicting other kinds of behavior.

In an earlier paper, we reported some preliminary work on the relationship of the interaction variables to such organismic variables as age and sex (9). In the present study, we chose to explore these relationships further and to ascertain also the relationship of interview behavior to certain psychological test measures which are currently in clinical use, especially to those test variables which, on the surface, would appear to be measuring similar aspects of interpersonal behavior. For example, we believed that certain characteristics of an individual's verbal interview behavior could well be related to Wechsler Verbal IQ or to socioeconomic level, inasmuch as higher educational background and higher occupational status presumably are associated with the acquisition and continued use of certain verbal and social skills. We also expected that Rorschach variables might correlate with some aspects of interview behavior as measured by the Interaction Chronograph. Rorschach reaction-time, for instance, presumed to reflect habitual speed of response or, in Beck's words, "lack of inhibition" (2, p. 52), might be positively correlated with length of delay before making a verbal or other communicative response to the interviewer; and Rorschach Total Time and R (total number of Rorschach responses) might be related to the amount of interview time in which the interviewee was speaking. In other words, both situations (Rorschach and interview) might be found to measure, among other things, a characteristic readiness or quickness to respond, and a characteristic level of activity or productivity.

Selection of the test variables was of course based upon such hunches or expectations as those mentioned above. There was, however, no systematic attempt to construct a series of hypotheses since we did not want to feel committed to a particular theoretical viewpoint when our knowledge about the "meaning" of Interaction Chronograph variables is still sketchy and tenuous. Rather, it was felt that an exploratory study was all that could be attempted in the preliminary stages of validating these interviewee interaction patterns, and that any correlations which were found in the initial study of 20 *S*s could serve as hypotheses for a cross-validation or replication study with a second group of 20 *S*s.

METHOD

The 40 *S*s in the original and cross-validation group were 26 outpatients and 14 inpatients (all white) of the Massachusetts General Hospital. The presenting problems of the patients, and the procedure of the interview were similar to those of other populations studied earlier (7, 12, 13). There were 25 females and 15 males ranging in age from 15 to 80, the mean age being 40. The same psychiatrist conducted all interviews, and the same observer recorded them from the other side of a large one-way mirror. A microphone was hung from the ceiling, and the voices were amplified and conducted into earphones worn by the observer. Facilities for visual and auditory observation were excellent.

Fifty patients, altogether, were interviewed in the order of their referral to the interviewer for psychiatric evaluation, their selection being without known bias. After the interview, the patients were referred for psychological testing. The psychologist had no knowledge of the patient's interview interaction behavior in order to insure independence of the psychological test measures and interview observations. Of the 50 referrals, 10 had to be eliminated from the experimental series because of refusal to take the psychological tests, borderline intelligence, or severe speech handicap. The tests were administered in one sitting, in uniform order, and with uniform instructions. First, educational and occupational data were obtained for use in determining socioeconomic level on the Hollingshead Index of Social Position. On this scale (4) the patients ranged from Class 1 (high executives of large concerns, proprietors, major professionals, etc.) to Class 5 (unskilled employees, welfare recipients, etc.). The mean for the population was Class 4 (skilled and semi-skilled workers, etc.). The complete Wechsler Adult Intelligence Scale was administered (16), followed by the simple language version of the Taylor Manifest Anxiety Scale (14, p. 288) and the Saslow Psychosomatic Screening Inventory (11). The next instrument administered was the Rorschach. Lest the amount of examiner-patient verbal interaction vary and possibly affect the latter's productivity, the examiner gave standardized instructions and began the Rorschach without further comment.[3] If the patient rejected the first card or gave only one response, he was reminded to "be sure to report *everything* you see." Subsequent rejections were accepted without comment and questions were answered merely by a nod of the head or a word or two if necessary. This manner of presenting the Rorschach may have resulted in a smaller number of responses by some patients than would have been obtained with more encouragement. However, it was felt necessary to keep stimulus conditions, i.e., the Rorschach examiner's behavior, as similar as possible from patient to patient in order to make the Rorschach situation as standardized as were the interview conditions. The mean Rorschach total time of 8.23 minutes (shown in Table 1) was somewhat low, probably due in part to the manner of presentation and the nature of the population, but primarily to the fact that in this study total time was calculated as the time during which the patient was actually responding to the cards rather than as the total duration of the test. Last of all, the patient was asked to fill out the Allport Ascendance-Submission Scale. It was occasionally necessary, with those patients having the lower IQs, to help interpret items. It can be seen from the group mean for this test in Table 1 that the 40 patient *S*s obtained a score (−15.95), which presumably indicates greater submissiveness than the average college student, who obtains a mean score of approximately 0.

In order to control for interviewer (and psychological test examiner) experience differences, an odd-even method was used for selecting the 20 *S*s of the original study and the 20 *S*s of the cross-validational sample. Group 1 consisted of patients 1, 3, 5, etc., and Group 2 of the alternate patients. Statistical procedures were subsequently performed on each group separately, and then on both groups together.

TABLE 1

MEANS, STANDARD DEVIATIONS AND RANGES OF THE PSYCHOLOGICAL TEST AND ORGANISMIC VARIABLES FOR THE COMBINED SAMPLE OF 40 SUBJECTS

Variable	Mean	*SD*	Range
Age	40.33	13.93	15 to 80
Total IQ	97.65	12.70	77 to 129
Verbal IQ	101.45	13.40	77 to 137
Performance IQ	93.25	13.45	70 to 127
Taylor Anxiety	26.60	11.02	6 to 45
Psychosom. Inv.	13.50	10.14	0 to 44
Ascend.-Submiss.	−15.95	19.40	−47 to +39
Socioecon.			
(raw score)	58.40	13.92	11 to 77
(class)	4.12	1.12	1 to 5
Rorschach:			
Reaction time	17.98″	14.05″	4″ to 85″
Total time	8.23′	4.50′	1′ to 20′
No. Responses (R)	16.43	10.19	2 to 51
A + Ad%	56.18	16.56	24 to 99
No. of Categ.	6.58	4.48	2 to 13
% Popular	33.75	22.50	0 to 99
No. of Card Rejects	1.45	2.10	0 to 8

RESULTS

Table 1 contains the means, standard deviations, and ranges of the psychological test and organismic variables for the total sample. There were no statistically significant differences between the two groups for these (test and organismic) variables. Means, etc., for the interview measures were comparable to those of groups previously studied (7, 12, 13). Table 2 contains the correlations (Pearson *r*) between the 14 interview variables on the one hand, and 13 psychological test measures, plus age, sex, and Hollingshead Index of socioeconomic status, on the other. The original matrix of intercorrelations contained 35 organismic and psychological test variables, but 19 are not presented here because they were found to be unrelated to any of the interview variables and would consume an inordinate amount of space. The excluded measures consisted of the Allport Ascendance-Submission Scale (which, developed on a college population, may have been merely unsuitable to the present clinic population) and a number of Rorschach scoring determinants. The Beck system was used for scoring Rorschach protocols, the only new classification used being "number of categories." This referred to the number of different Rorschach content categories utilized by a particular patient and was included in the present study because the distribution of scores among Beck's separate content categories was especially diffuse, with low frequencies in any single content classification making inappropriate the application of conventional correlation techniques. The distributions of the other variables differed somewhat in the degree to which they fulfilled the assumptions of the Pearson *r* but, in general, did not deviate markedly from a normal distribution. It was felt that, in view of the above as well as the Norton findings quoted in Lindquist (6), little precision would be lost in using Pearson *r*'s on the data in their original form. This is es-

---

[3] The exact instructions were as follows: "I am going to show you some cards, and I want you to tell me whatever you see on them. This can involve the whole card or any part of it. Most people see a number of things on each card, so look carefully for everything you see represented there. When you are sure you have told me everything you see, put the card down and take the next one."

TABLE 2

PSYCHOLOGICAL TEST CORRELATES OF INTERVIEW INTERACTION BEHAVIORS FOR TWO GROUPS OF 20 *S*s EACH AND COMBINED TOTAL GROUP OF 40 *S*s

	Pt.'s Units			Pt.'s Action			Pt.'s Silence			Pt.'s Tempo		
	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.
Sex												
Age							−.56	−.45	−.49			
Total IQ	−.24	−.45	−.34				.27	.23	.26			
Verbal IQ	−.36	−.40	−.36	.20	.21	.20	.21	.29	.26	.22	.28	.25
Perf. IQ												
Anxiety							−.23	−.20	−.20			
Psychosom.	−.38	−.41	−.39	.29	.50	.39				.29	.52	.40
Socio-Econ.							−.20	−.31	−.22			
Ave. RT	−.39	−.08	−.25	.60	.20	.44				.60	.17	.42
Tot. Time							−.27	−.23	−.25			
No. Responses												
D%	.39	.11	.25	−.42	−.26	−.35				−.43	−.30	−.36
Dd%	−.60	−.37	−.47	.65	.25	.48				.65	.29	.48
A + Ad%	.41	.24	.31	−.45	−.13	−.29				−.43	−.14	−.28
% Popular												
No. Categ.												

	Pt.'s Activity			Pt.'s Adjustmt.			Dr.'s Adjustmt.			Pt.'s Synchron.			Dr.'s Units		
	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.
Sex															
Age				.24	.30	.27	.31	.31	.30				−.64	−.11	−.31
Total IQ							−.37	−.41	−.36				−.25	−.44	−.33
Verbal IQ	.18	.24	.21				−.42	−.35	−.34				−.38	−.38	−.36
Perf. IQ							−.23	−.41	−.31						
Anxiety										−.28	−.29	−.28			
Psychosom.	.29	.52	.40							−.36	−.43	−.39	−.34	−.43	−.38
Socio-Econ.							.31	.26	.23						
Ave. RT	.60	.16	.44							−.58	−.12	−.38			
Tot. Time															
No. Responses															
D%	−.42	−.31	−.37							.45	.28	.37	.46	.16	.31
Dd%	.65	.32	.51							−.53	−.31	−.43	−.61	−.37	−.48
A + Ad%	−.46	−.17	−.32							.47	.21	.33	.46	.25	.34
% Popular															
No. Categ.															

	Pt.'s Initiative			Pt.'s Quickness			Pt.'s Dominance			Pt.'s No. of Interrup.			Length Interview		
	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.
Sex							.41^{e}	.58^{e}	.50^{e}	.04^{e}	.63^{e}	.36^{e}			
Age	.50	.45	.47	.48	.45	.47									
Total IQ							−.11	−.54	−.33	−.18	−.41	−.32	.13	.33	.24
Verbal IQ							−.23	−.61	−.42	−.29	−.45	−.38	.28	.27	.27
Perf. IQ															
Anxiety	.57	.36	.44	.51	.25	.37									
Psychosom.	.28	.28	.27										.25	.52	.39
Socioecon.															
Ave. RT										−.33	−.16	−.21	.59	.20	.40
Tot. Time	.35	.24	.31	.25	.36	.31	−.26	−.33	−.27	−.25	−.22	−.20			
No. of Responses							−.36	−.43	−.37	−.22	−.21	−.18			
D%							.25	.23	.23	.28	.28	.24	−.34	−.38	−.33
Dd%							−.28	−.27	−.25	−.35	−.26	−.24	.54	.26	.37
A + Ad%				−.32	−.27	−.29							−.50	−.07	−.26
% Popular				−.41	−.15	−.29	.37	.39	.37						
No. of Categ.							−.25	−.38	−.33						

	$N = 20$	$N = 40$
Note:—r required for significance at .05 level is	.44	.31.
r required for significance at .01 level is	.56	.40.

pecially true since, in previous studies of interview behavior (9), comparison of Pearson r's and rho's have yielded negligible discrepancies in most cases.

Table 2 contains only those correlations which are significant (p = .01 or .05) in both the original and cross-validational groups or which indicate for both groups a consistent trend approximating significance. Thus, Table 2 includes only those correlations which appear in group one and are replicated in the second group of 20 since, in a matrix of 35 by 14 variables, one would expect to find a number of misleading "significant" correlations based on chance alone.[4]

While all fourteen interview variables are presented in Table 2, the present paper will focus on only five major variables because most of the others are some kind of combination or derivation of these. The first and second of the five variables to be discussed are Pt.'s Units, (the number of times the patient talks or communicates by gesture during the standardized interview) and Pt.'s Action (the average duration of his utterances). As reported in a previous paper (9), Pt.'s Units has a high negative correlation with Pt.'s Action. (In the present total sample, this correlation is −.85)[5]. In Table 2 it can be seen that the correlation coefficients between the test variables and the interaction variables also seem to define these two interview characteristics as opposites, and thus begin to give further definition to them. For example, in the two samples combined, the negative r's between Pt.'s Units and Total IQ (−.34) and Verbal IQ (−.36) indicate that many short interview units are significantly associated with lower Total and Verbal IQ, while fewer utterances are significantly associated with relatively higher Total and Verbal IQ. The positive (but not statistically significant) r's of .20 and .21 between Pt.'s Action and Verbal IQ may indicate a tendency in both groups for the longer utterances to be found in those with the higher Verbal IQ. Taking these observations together, it would appear that during the interview the more intelligent patients tend to speak a smaller number of times but with longer durations per utterance, while the less intelligent patients speak a greater number of times but with a shorter duration each time.

Table 2 presents other psychological test correlates of the number of times (units) a patient interacts during the standardized interview. Thus the r of −.25 between Pt.'s Units and Average Reaction Time; of .25 between Pt.'s Units and D%; of −.47 between Pt.'s Units and Dd%; and of .31 between Pt.'s Units and A + Ad%, describe the (less intelligent) patient with many short interview units as responding more quickly to the Rorschach cards, tending to give more large detail responses but fewer small detail and more animal responses. In parallel fashion, the correlation coefficients between Pt.'s Action and the same Rorschach variables portray the (more intelligent) patient with the longer interview utterances (and consequently a smaller number of them) as responding more slowly to each Rorschach card, tending to give fewer large detail but more small detail responses, and fewer animal responses. However, intelligence level alone does not account for the above observations since, in the two samples studied individually or combined, the correlation coefficients between IQ and D% and IQ and Dd% were not statistically significant. (These coefficients are omitted from Table 2).

The r's of −.39 between Pt.'s Units and Psychosom. (the number of symptoms checked by the patient on the Saslow Screening Inventory) and .39 between Pt.'s Action and Psychosom. indicate that the patient who talks a greater number of times but with short utterances in the interview reports fewer psychosomatic symptoms in relation to stress in his daily life. (There was no significant association between IQ and number of symptoms checked on the inventory). The explanation of these r's of −.39 and .39 is unclear. Perhaps they reflect denial, caution, or repression on the part of these patients of relatively lower IQ, who tend also to give more stereotyped Rorschach responses and who speak many times in the interview because, under the conditions of the standardized interview, each of their utterances is brief.

---

[4] It is clear in Table 2 that many of the correlation coefficients between interview and psychological test variables, while reaching the .05 or .01 levels of statistical significance are, nevertheless, of a low order of magnitude. Taken together, however, they seem to contribute to the understanding of interview behavior.

[5] This is partly a function of the partially standardized interview. That is, three of the subperiods are fixed in duration, and thus the patient who speaks in long utterances will not have the opportunity to use a large number of units and conversely, the individual with a relatively large number of units must have spoken in fairly short utterances.

Certain other significant correlates of Pt.'s Units and Pt.'s Action (not shown in Table 2) are of interest. These are *r*'s of −.33 between Pt.'s Units and Initiative; of .34 between Pt.'s Units and Pt.'s Dominance; and of .56 between Pt.'s Units and Pt.'s No. of Interruptions. The corresponding *r*'s with Pt.'s Action are .29, −.38, and —.62, respectively. These coefficients portray the patient who talks often during the interview as taking less initiative in the "silence stress" subperiod, and as interrupting and talking down the interviewer more often during the "interruption stress" subperiod than the patient who speaks a smaller number of times but in longer units.

In summary, the person with a large number of Pt.'s Units and the person with few units and correspondingly longer durations of utterance per unit present almost opposite pictures. In Table 2, there is a consistent though nonsignificant trend in both groups for the person with longer utterances to be more intelligent (*r* of .20 between Pt.'s Action and Verbal IQ). He responds more slowly to the Rorschach (*r* of .44 between Pt.'s Action and Average RT), but apparently with more individually meaningful material (*r* of .48 with Dd%) in the sense that he reports seeing more rare details, and he gives fewer of the relatively easy large detail (*r* of −.35 with D%) and animal responses (*r* of −.29 with A + Ad%). He also reports having more psychosomatic symptoms, as reflected in the Saslow Screening Inventory (*r* of .39 with Psychosom.) and for this reason might be thought of as being less guarded in his test-taking behavior. In line with this hypothesis, he is again the opposite of the "many units but each of short duration" person in that he tends to show more initiative (*r* of .29 between Pt.'s Action and Pt.'s Initiative, not shown in Table 2) in the ambiguous part of the interview when the interviewer remains silent by consistently failing to respond to the patient. Finally, the more intelligent, "few units but each of long duration," individual lacks the less intelligent "many units but each of short duration" individual's tendency to dominate and interrupt (*r* of −.38 between Pt.'s Action and Pt.'s Dominance, and of −.62 with Number of Interruptions). Pt.'s Initiative and Pt.'s Dominance, themselves, are not significantly intercorrelated. These figures are not shown in Table 2.

It is interesting to note in Table 2 that the average duration of Pt.'s Silence is *not* the opposite of Pt.'s Action in regard to the psychological test correlates. This finding is further substantiated by a recently completed factor-analysis in which it was found that these two interview variables (average Pt.'s Action and average Pt.'s Silence) are almost wholly independent factors. In fact, the results in Table 2 indicate that intelligence (both Total and Verbal) is positively related to Pt.'s Silence just as to Pt.'s Action. The relationship falls just short of statistical significance, but is consistent in both groups between Verbal IQ and both Pt.'s Action and Pt.'s Silence, and additionally between Total IQ and Pt.'s Silence. Another consistent trend is for Pt.'s Silence to be positively correlated with higher socioeconomic level. (The Hollingshead Index ranges from a high score for the lowest socioeconomic group to a low score for the highest socioeconomic group, thus the negative correlation.) Pt.'s Silence is also negatively associated with Taylor Anxiety score. It thus would appear that silence behavior in the interview has a low positive relationship to intelligence and socioeconomic level and is negatively related to anxiety, with the longer average interview silences tending to be found in patients of higher intelligence, higher socioeconomic level, and lower Taylor anxiety levels. In Table 2 it can be seen further that the patient's age is the variable having the greatest relationship to Pt.'s Silence (*r* of −.49), indicating that younger *S*s had the longer average interview silences.

On the basis of these results, one might wonder if Pt.'s Silence behavior has two aspects: one, a degree of tolerance for silence among patients with lower anxiety scores and also in patients of higher socioeconomic status and higher IQ; the other, a factor associated with age (poise, self-assurance, or lack of defensiveness) and resulting in longer waiting periods before taking the initiative in the silence stress subperiod among the younger *S*s. The former culturally "positive" aspects of Silence behavior account, however, for only a small proportion of the total variance, the correlations being consistent in both the original and cross-validational samples but not attaining statistical significance. The latter "negative" and larger aspect of Pt.'s Silence behavior seems to be a lack of ability in the younger *S*s to respond as quickly (with as short silences) as the older *S*s in the silence stress subperiod of the interviews. In this regard it should be mentioned (data not presented in Table 2) that Pt.'s Silence and Pt.'s Initiative have a correlation of −.69 in this sample, which indicates that whatever accounts for the Pt.'s Silence variable (an average duration measure) partly accounts also for the *S*'s failure to take the initiative after a double silence in Period 2 (a frequency measure).

Turning to the variable, Pt.'s Initiative, also shown in Table 2, it can be seen that age has both a correlation of −.49 with Pt.'s Silence (the older the patient the shorter his average interview silences) and +.47 with Pt.'s Initiative (the older the patient the more often he took the initiative during the silence period), thus supporting the possibility of the presence of a "poise" factor, increasing with age. Further suggestion of the plausibility of this "poise" interpretation is given by a recent analysis of data from a previously studied population of 60 out-patients (9). These results are shown in Table 3. For these 60 *S*s, the mean initiative score of those under age 30 was .65 during the first interview, while the mean initiative score of those *S*s over 30 was .78 (*F* test significant at .01 level). The value of the correlation ratio, epsilon, was .34 (significant at .01 level), indicating that this relationship between age and increasing interview initiative is a moderate one. All of the *S*s had a second interview, at which time the age-initiative relationship disappeared; mean initiative scores for the younger and older groups were .70 and .73 respectively, with the *F* test between age and initiative dropping to an insignificant .39 and epsilon to zero. This finding of less initiative among the younger group in the first interview is cross-validated in the present sample of 40 *S*s. The epsilon value of .46, shown in Table 3, corresponds to the Pearson *r* of .47 (for the same data) shown in Table 2.[6]

TABLE 3

PATIENTS' AGE AND INITIATIVE DURING INTERVIEW

Mean Initiative Score	Patient's Age		*F* test	*e*
	Under 30	Over 30		
Present sample, $N = 40$				
a. First interviews	.56	.77	11.88**	.46**
Previous sample, $N = 60$				
a. First interviews	.65	.78	7.71**	.34**
b. Second interviews	.70	.73	.39	.00

** Significant at the .01 level of probability.

Given the hypothesis of a "poise" or "set" factor, developing with age, it appears to operate largely in the first interview and apparently produces no behavioral effect when the patient has some familiarity with the stimulus situation, as in the second or retest interview. This hypothesis has certain face validity when one thinks of either therapeutic or social contacts with adolescents, who seem to have considerable reluctance to initiate with adults but who manage to respond at some length when the initiative has been taken by someone else. At any rate, it is interesting that age is related to Pt.'s Initiative and Pt.'s Silence, but not to Pt.'s Action.

Other psychological test correlates of Pt.'s Initiative, presented in Table 2, further suggest that it is, in some ways, the opposite of the Pt.'s Silence variable. That is, Pt.'s Initiative increases positively with anxiety score and number of psychosomatic symptoms, just as Pt.'s Silence is negatively related to anxiety. Thus, one could speculate that Pt.'s Initiative is, in part, a function of the magnitude of his anxiety level as the latter is assessed by the Taylor Anxiety scale; the more anxious *S*s show the higher percentage of interview initiative (*r* of .44). That initiative may be a reflection of some kind of "drive" ("anxiety") is further suggested by the fact that it seems to be related to effort expended on the Rorschach, since it is significantly and positively correlated ($r = .31$) with total time spent in responding to the ink blots. Possibly the percentage of times an *S* takes the initiative under conditions of planned interview silence may be correlated with a variety of laboratory learning tasks, as in the Taylor "anxiety" level.

In Table 2 it can be seen that Pt.'s Domi-

---

[6] Twenty of this group have been re-interviewed after an 8-month interval. For these 20, the age-initiative relationship again disappears in the second interview. The mean initiative score for the under 30 group is .66, while for the over 30 group it is .72. The *F* test of .67 is not significant; nor are the *r* of −.09 and the *e* of .00. The same 20 *S*s showed a positive (*r* of .26) age-initiative relationship in their first interviews.

TABLE 4

PATIENTS' SEX AND DOMINANCE-SUBMISSION PATTERN DURING INTERVIEW

Mean Dominance Score	Females	Males	*F*-test	ϵ
Present sample, $N = 40$				
a. First interviews	−.42	−.71	8.44**	.50**
Previous sample, $N = 60$				
a. First interviews	−.33	−.49	5.65*	.26*
b. Second interviews	−.26	−.53	12.50***	.40**

* Significant at the .05 level of probability.

** Significant at the .01 level of probability.

*** Significant at the .001 level of probability.

nance seems to be the only variable which is significantly related to sex, the female patients being relatively less submissive (or more dominant) than males with the male interviewer. Since this dichotomous variable (sex) was not suited to Pearson *r*'s, *F* tests and epsilon correlations were obtained and are the values shown (.41, .58, and .50) across the sex row in Table 2. The values of epsilon needed for significance are somewhat lower than those for *r*, being .39 and .53 for the .05 and .01 levels, respectively, with an *N* of 20, and .27 and .37 with an *N* of 40.

A further analysis of the sex-dominance relationship is presented in Table 4. Mean Pt.'s Dominance score for females in the present combined sample of 40 *S*s was found to be −.42 and for males, −.71. The *F* value for this difference was 8.44, significant at the .01 level, as was the epsilon correlation ratio of .50. Since this finding of significantly less submissiveness in the female patients relative to the male patients was contradictory to expectations, built upon cultural stereotypes of the dominant male, one may wonder if the factor of illness is not an important one here. Perhaps one correlate of these persons' psychiatric difficulties is that they do not fit into their expected cultural roles. We hope to repeat the sex-dominance analysis on a "normal" group to test the validity of this hypothesis. However, the finding does appear to be consistent among patient groups since the same relationship was obtained in a previously studied sample (9). As shown in Table 4, in that sample of 60 outpatients, the epsilon correlation value was .26, significant at the .05 level of confidence for the first interviews. For the second interviews, the epsilon correlation increased to .40, significant at the .01 level. The *F* tests similarly indicate an increase in this relationship for the second interview. Thus, in both the first and retest interviews, the males were found to be more submissive (were dominated more) than were the females. However, the negative mean dominance scores indicate that both male and female patients, as groups, are more submissive than dominant, with male patients being more so. Since we have not yet studied a group of normal subjects of comparable characteristics, we do not know if this difference exists only in a patient population and if so, whether women patients are relatively more dominant than the normal members of their sex, whether the males are more submissive than the normal members of their sex, or both.

Returning to Table 2, we see further that Pt.'s Dominance was negatively associated with Total and Verbal IQ, i.e., the less intelligent the *S* the more dominance and interruption he showed in the interview. Looking to the Rorschach correlates of the dominance variable, we find a picture of stereotypy or guardedness. Patients high in Pt.'s Dominance tended to spend less total time with the ink blots, gave fewer responses, reported significantly more large detail responses, perceived more populars, attended to fewer small details, and used a smaller number of content categories.

Thus, the patient who shows relatively more interview dominance tends to be less intelligent, displays a relative lack of imagination on the Rorschach test, and apparently either has less ego-involvement in his Rorschach performance or is significantly more guarded about his responses, returning cards relatively quickly and after fewer responses. He spends less total time in Rorschach performance, although his reaction time does not differ from those who are less dominant, so that he apparently does not respond hastily and without thinking but simply differs in the quantity (minus *r* with Total Number of Rorschach responses) and quality (plus *r* with P%, D%, and minus *r* with No. Categ.) of his productions. This relative stereotypy or rigidity of the high interview Dominance individual is another interesting finding which merits further exploration.

It may be noted that throughout the analysis the correlations between several Rorschach variables and the Interaction Chronograph variables are lower in Group 2 than in Group 1. Inspection of the sigmas of these variables (not shown separately in Table 1) revealed that, in the case of reaction-time, the sigma of Group 2 was less than half that of Group 1, while for D% and Dd% the sigmas were ⅔ the size of those for Group 1. Since the size of the Pearson *r* is dependent upon variability within the group or is lowered by a restricted range of scores, the smaller standard deviations may account for the lower correlations in Group 2. The variable A + Ad%, however, (which also had a lower correlation with the interview variables in Group 2) did not present a similar difference in range and sigma between the two groups, and this factor therefore does not explain the discrepancy in the case of A + Ad%.

DISCUSSION

The present observations suggest the possibility of differentiating among individual patients along several definable dimensions. There seem to be some consistent organismic and psychological test differences between people who speak in long and short utterances in the partially standardized interview, between those who have short or long silences, between those who show much or little interview "initiative" in the silence stress subperiod, and between those who "submit to" or who "dominate" the interviewer when the latter has interrupted them in the interruption stress subperiod.

Intelligence was found to be associated with certain interview interaction patterns, although socioeconomic level was not. It should be noted, however, that in our combined clinic and inpatient population, we were not able to obtain an adequate sampling of the higher socioeconomic levels and do not feel that we have explored adequately the possibility of a relationship between socioeconomic level and interview interaction pattern.

In concurrence with the negative results of several other recent investigators, in this study Rorschach perceptual determinants were not found to correlate with interview behavioral characteristics, although a few Rorschach content measures proved to be more promising. This failure of such Rorschach variables as Movement, Color, etc., to correlate with interview behavior is in some ways similar to the recent findings of Knopf (5), who reported that a few Rorschach content categories were useful in differentiating among widely differing diagnostic groups, while Rorschach perceptual variables failed even to differentiate reliably a schizophrenic group from other diagnostic groups.

While the present investigation was designed to study the possible psychological test correlates of interview interaction behavior, and thus to give meaning to individual differences in the interview, it may be observed that the interaction variables may themselves be thought of as contributing meaning to the psychological test variables with which they are significantly associated. That is, the standardized interview could equally well have been chosen as the criterion against which to validate the psychological tests. For example, we have shown that those *S*s of higher intelligence have different interview characteristics than those of lower intelligence, since they speak in longer utterances and tend to interrupt or dominate the interviewer less frequently. Similarly, it is interesting that those *S*s who give stereotyped Rorschach behavior in terms of fewer responses, more large details, more populars, fewer content categories, etc., tend to dominate the interviewer and can, in that respect, be said to show somewhat less social sensitivity in their interview interaction. Also, those who score relatively high on the Taylor Anxiety scale are less silent and take the initiative more frequently than those who are presumably less anxious. A parallel could be drawn between this and some of the recent literature indicating that those with higher anxiety scores tend to obtain higher learning scores on conditioning or easy learning tasks.

The present study, in any event, is to be considered only an initial approach to an understanding of the validity or "meaning" of the interview behaviors measured by the Interaction Chronograph. It has been an exploratory approach through which we hoped to gain leads for more systematic research. The examination of psychological test correlates has been especially rewarding when one considers that in the present study we have related them only to the simplest, discrete interview variables, and so far have not attempted to relate them to patterns of these variables or to sequential changes in behavior as the interview proceeds.

The relationship of the psychological test and other variables to more complex interview interrelationships and to sequential variations in interviewee behavior demands systematic inquiry.

Summary

The present study was an attempt to examine the "meaning" of Interaction Chronograph patterns through concurrent validity, relating interview interaction patterns and psychological test behavior. A number of meaningful correlations was found in a sample of 20 patient *S*s and cross-validated in a second similar sample. Individuals who speak in short utterances during the interview are less intelligent and seem to be more "on their guard" or more stereotyped in their psychological test productions. Those with long utterances are somewhat more intelligent and are less cautious or denying, giving less conventional performance on the Rorschach test and reporting more psychosomatic symptoms. Initiative during the silence period of the interview is hypothesized to be related to "poise" or self-assurance (or some other "set" factor) since it seems, in the first of a pair of interviews, to be greater with age. The individual with a high Pt.'s Dominance score is of relatively lower intelligence and presents a Rorschach picture of a smaller number of responses, high P% and A + Ad%, a smaller number of content categories, etc. This was interpreted as indicative of stereotypy or rigidity. Pt.'s Dominance seems also to be the only interview variable which is related to sex, female patients being relatively less submissive than male patients. The present investigation can also be thought of as providing certain behavioral (interview) correlates of psychological tests (Wechsler, Rorschach, etc.) currently in clinical use.

REFERENCES

1. Allport, G. W., & Allport, F. H. *The A-S Reaction Study: A scale for measuring ascendance-submission in personality.* (2nd ed.) Boston: Houghton Mifflin, 1939.
2. Beck, S. J. *Rorschach's test—a variety of personality pictures.* Vol. 2, New York: Grune & Stratton, 1949.
3. Chapple, E. D. The Interaction Chronograph; its evolution and present application. *Personnel*, 1949, **25**, 295–307.
4. Hollingshead, A. B., & Redlich, F. C. *Social class & mental illness.* New York: Wiley, in press.
5. Knopf, I. J. Rorschach summary scores in differential diagnosis. *J. consult. Psychol.*, 1956, **20**, 99–104.
6. Lindquist, E. F. *Design and analysis of experiments in psychology & education.* Cambridge: Riverside Press, 1953.
7. Matarazzo, J. D., Saslow, G., & Guze, S. B. Stability of interaction patterns during interviews: a replication. *J. consult. Psychol.*, 1956, **20**, 267–274.
8. Matarazzo, J. D., Saslow, G., & Matarazzo, Ruth G. The Interaction Chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347–367.
9. Matarazzo, J. D., Saslow, G., Matarazzo, Ruth G., & Phillips, Jeanne S. Stability and modifiability of personality patterns during a standardized interview. In P. A. Hoch and J. Zubin (Eds.), *Psychopathology of communication.* New York: Grune & Stratton, 1957.
10. Phillips, Jeanne, S., Matarazzo, J. D., Matarazzo, Ruth G., & Saslow, G. Observer reliability of interaction patterns during interviews. *J. consult. Psychol.*, 1957, **21**, 269–275.
11. Saslow, G., Counts, R., & DuBois, P. H. Evaluation of a new psychiatric screening test. *Psychosom. Med.*, 1951, **13**, 242–253.
12. Saslow, G., Matarazzo, J. D., & Guze, S. B. The stability of Interaction Chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, 1955, **19**, 417–430.
13. Saslow, G., Matarazzo, J. D., Phillips, Jeanne S., & Matarazzo, Ruth G. Test-retest stability of interaction patterns during interviews conducted one week apart. *J. abnorm. soc. Psychol.*, 1957, **54**, 295–302.
14. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
15. Ulett, G. A., Gleser, Goldine, Starr, P., Haddock, J., Lingley, L., & Lawler, Ann. Psychiatric screening of flying personnel: Further studies toward the development of an electroencephalographic screening technique. *USAF Sch. of Aviat. Med.*, 1953, Project No. 21-0202-0007 (Report No. 5).
16. Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corp., 1955.

Received May 28, 1957.

INTERRELATIONSHIPS AMONG INDIVIDUAL DIFFERENCE VARIABLES, BEHAVIOR IN PSYCHOTHERAPY, AND VERBAL CONDITIONING[1, 2]

IRWIN G. SARASON
University of Washington

THE present experiment dealt with the interrelationships among three variables: (*a*) *S*'s score on the scales of a personality inventory, (*b*) therapists' ratings of *S*'s behavior in the psychotherapy situation, and (*c*) *S*'s performance in a verbal conditioning situation.

Within recent years there has been an increase in the frequency of studies on verbal conditioning (1, 4, 9, 13, 14). These studies have centered on the effects of *E*'s reinforcement of *S*'s verbal responses (of a predetermined class) on *S*'s subsequent verbal behavior. One is struck by the similarity of this two-person verbal conditioning situation to psychotherapy and everyday conversational situations. In both of these kinds of interpersonal interactions, individuals emit responses that may or may not be reinforced by another individual. For example, in the psychotherapy situation, the therapist may be particularly interested in reinforcing responses (verbalizations) of the shy withdrawn patient which indicate assertiveness and independence. On the other hand, with other kinds of patients and at certain stages of psychotherapy, the therapist may be inclined to reinforce responses of the patient which indicate dependence on or compliance with the therapist.

Implicit in this example is the importance of initial individual differences in *S*s' anticipation of reinforcement and their reaction to reinforcement. In a random sample of *S*s, one would not expect all *S*s to respond to reinforcement in the same way. Tatz (10), for instance, has recently demonstrated the significant role of individual differences among *S*s in a verbal conditioning situation in the extent to which they anticipate punishment or reward from the experimenter. The present study was concerned with the effect of certain personality characteristics on verbal conditioning.

The instrument employed to select *S*s was the Autobiographical Survey, a true-false test that yields six scores: (*a*) Test Anxiety (TA), (*b*) General Anxiety (GA), (*c*) Lack of Protection (LP), (*d*) Need for Achievement (NA), (*e*) Defensiveness (D), and (*f*) Hostility (H). Brief descriptions of the scales are presented below. While the names of five of the six scales fairly well suggest the nature of their content, it should be mentioned that the Lack of Protection scale consists of items all of which were written to adhere to Freud's definition of anxiety, i.e., feeling unprotected and unable to cope with novel, threatening situations.

Two predictions were made about the verbal conditioning performance of *S*s differing in scores on certain scales: 1. *S*s with high scores on the Test Anxiety (TA), General Anxiety (GA), and Lack of Protection (LP) scales condition faster than *S*s with low scores on these scales. These predictions were based on the hypothesis that in a novel situation such as verbal conditioning, in which *S* has no prior knowledge of what constitutes a correct response, anxious individuals, i.e., individuals who admit fear, tension, and inadequacy feelings in strange situations, are relatively more likely than other *S*s to increase the frequency of responses reinforced by *E*'s uttering "mm-hm." Too, individuals who feel inadequate in coping with strange situations may be more likely to accept direction (in this case, *E*'s reinforcement) from an authority figure. 2. Individuals with high scores on the Defensiveness and Hostility scales condition significantly more slowly than low scoring *S*s. This prediction was based on the hypothesis that people who are defensive or hostile in interpersonal situations may frequently be unlikely to accept interpretations, reinforcements, and cues from others.

At this point it is well to emphasize that the predictions made for the verbal conditioning

---

[1] This research was carried out at the West Haven, Connecticut VA Hospital. The writer is indebted to Jacob Levine, chief psychologist, for his cooperation and encouragement while the research was in progress. John Rakusin contributed many helpful suggestions, for which the writer is grateful.

[2] Portions of this paper were presented at the May, 1957, meeting of the Western Psychological Association.

behavior of *S*s are similar to a set of predictions made concerning the behavior of patients in psychotherapy. In carrying out this study, attention was focused on the possibility of studying complex variables of presumed importance in psychotherapy (e.g., the defensiveness of the patient) in simpler, more controllable situations such as verbal conditioning. With a better understanding of the determinants of behavior in such relatively simple interpersonal situations, it is possible that lines of experimental attack may be opened in more complex situations such as psychotherapy. In order to gain some preliminary information on this possibility, therapists' ratings of patients' behavior in therapy were obtained for the majority of *S*s in this experiment. Thus, it was hoped that some initial rough comparative statements might be made concerning the behavior of *S*s in both the therapeutic and verbal conditioning situations.

The *S*s were rated by their therapists on three characteristics which seemed to be related to certain characteristics of *S*'s Autobiographical Survey performance. These ratings concerned the degree of compliance with which the patient accepts therapist's interpretations, the degree of dependence the patient evinces towards the therapist, and the degree of hostility the patient displays towards the therapist. As with the H scale of the Survey, it was expected that *S*s rated by therapists as being hostile would condition more slowly than other *S*s. Similarly, high compliance and dependence therapist ratings were expected to be related to faster verbal conditioning on the basis of the need of the patient to please authority figures.

METHOD

Subjects

The *S*s were 60 neurotic and psychotic patients on the intensive treatment NP wards of the West Haven, Connecticut, VA Hospital. All but two of the *S*s were men, and the age range for the entire group was 22–52. All *S*s performed in the verbal conditioning situation and filled out the individually administered Autobiographical Survey.

Autobiographical Survey

This inventory consists of 180 statements to be answered True or False.[3] The first two items are neutral, nonscored items, and the remaining 178 are divided among the six scales. A brief description of each scale follows, together with examples of the kinds of items on the scale.

1. *Test Anxiety* (*TA*). This scale consists of 21 statements, most of which were taken from Sarason's Test Anxiety Questionnaire (6) and rewritten to fit the requirements of a true-false test rather than a graphic rating scale. Example: "While taking an important examination, I perspire a great deal."

2. *General Anxiety* (*GA*). The 24 items of this scale purport to tap the extent to which the individual experiences anxiety in a wide variety of situations other than testing. The statements used on this scale include some items to be found in the MMPI (5), the Taylor Anxiety Scale (11), and Sarason's General Anxiety Questionnaire (3). Example: "I worry about my social adjustment more than do most people."

3. *Lack of Protection* (*LP*). Some of the 42 items on this scale were taken from the scale presented by Mark (7). Most of them, however, were written by the author so as to adhere to the Freudian definition of anxiety: "The danger situation is the recognized, remembered and anticipated situation of helplessness. Anxiety is the original reaction to helplessness in the traumatic situation, which is later reproduced as a call for help in the danger situation" (2, p. 115) "... the situation of privation in which stimuli reach an unpleasurable magnitude and intensity without an ability to cope with them psychically and thus provide for their discharge, must represent to the infant a situation analogous to the birth experience, a repetition of the danger situation" (2, p. 76). Example: "When I was a child, I often wondered about how much my father loved me."

4. *Defensiveness* (*D*). Most of the 27 items of this scale were drawn from the K scale of the MMPI (5). This scale was included in order to obtain a measure of the individual's need to present himself in a favorable light. Example: "I am always patient with others."

5. *Hostility* (*H*). This scale, consisting of 23 items, taps the degree to which the individual admits to feelings of anger and hostility in a variety of situations. Many of these items were drawn from the MMPI. Example: "I rebel against bossy teachers."

6. *Need for Achievement* (*NA*). The 41 items of this scale, many drawn from the MMPI, were intended to measure the extent to which the individual feels called upon to attain a high level of performance in a variety of situations. Example: "My parents expected the highest level of achievement from their children."

Psychotherapists' Ratings

The therapists who made ratings of patients' behavior in psychotherapy sessions were chiefly psychiatric residents with one year's experience in psychotherapy. Some of the patients used in this study were rated by either staff clinical psychologists or advanced clinical psychology trainees, when these clinicians were the patients' therapists.

---

[3] A copy of the Autobiographical Survey and its scoring keys has been deposited with the American Documentation Institute. Order Document No. 5517, remitting $1.25 for 35-mm. microfilm or $1.25 for 6 by 8 in. photocopies.

The therapists were asked to rate the three behaviors in question as occurring *almost always*, *frequently*, *seldom*, or *never* in psychotherapy. The statements rated by therapists in the experiment were: (*a*) How compliant is this patient to the therapist's requests and in accepting his interpretations? (*b*) How dependent is this patient on the therapist? (*c*) To what extent does this patient express hostility to his therapist?

Procedure

The relationship in time for each *S* between administration of the Survey and performance in the verbal conditioning situation was not rigidly controlled because of exigencies in hospital schedule and routine. However, in the great majority of cases, *S*s were not administered the Survey until after they had performed in the verbal conditioning situation.

All *S*s were given the following preliminary instructions: "I am going to show you these cards one at a time. For each card make up a two-word sentence using the pronoun at the top of the card and one of the three verbs at the bottom. You may choose any verb you like for each card."

Following these instructions, *S*s received 160 verbal conditioning trials. The *E* did not reinforce any responses made by *S* on Trials 1–20. Each trial consisted of the presentation to *S* of a 3 x 5 card, on the top of which was typed the word *he* and on the bottom of which were typed three verbs of equal frequency of occurrence according to the Thorndike-Lorge lists (12). One of the three verbs was a verb of the reinforced class; the other two were neutral words not of the reinforced class. The reinforced class and the general procedure used were the same as those used in a recent study by Sarason (8). Reference should be made to that study for a detailed description of the verbal conditioning procedure. Verbs of the class Verbal Activity were chosen as the reinforced class, and whenever *S* used a verb of this class, *E* uttered "mm-hm." Verbal Activity included verbs having to do with the use of the vocal apparatus, e.g., *talked*, *whispered*.

Following the conditioning series, a brief interview was conducted along the lines described by Sarason (8) in order to determine the level of *S*'s ability to verbalize the basis upon which *E* reinforced *S* and, indeed, if *S* was even aware of *E*'s reinforcements. None of the *S*s as judged by the interviews was aware of the principle on the basis of which *E* emitted "mm-hm." Seven *S*s, however, did comment that they had noticed that *E* often uttered "mm-hm."

Results

Psychometric Findings

In order to assess meaningfully the influence on verbal conditioning of the individual differences tapped by the six personality scales of the Autobiographical Survey, it was of interest to obtain the intercorrelations as well as the reliabilities of the scales. Reliabilities, obtained by retesting 40 of the *S*s in this study two weeks after they initially took the Survey, were all above .70. Only the reliability of the NA scale fell below $r = .78$. The high interscale correlations in Table 1 suggest that the personality characteristics which the scales purport to measure are either defined in such a way as to be overlapping or are not "in nature" independent of each other. Whichever alternative is the case, however, in many of these correlations a large part of the variance in one variable is still unaccounted for by the variance in the other. For example, the TA and GA scales are correlated .46. This would suggest, for example, that although some *S*s admit to anxiety both in testing and a variety

TABLE 1

Intercorrelations Among Scales of Autobiographical Survey

($N = 60$)

	GA	TA	LP	NA	D	H
GA		.46	.40	.24	−.61	.37
TA			.36	.39	−.54	.36
LP				.44	−.52	.56
NA					−.31	.41
D						−.61
H						

$r = .28 = .05$ level.
$r = .33 = .01$ level.

of other situations, many individuals admit to anxiety in testing but *not* in many everyday life situations. Such findings suggest that in addition to pursuing the problem from the viewpoint of clarifying the concepts and factorially purifying the scales of the Survey, an important aspect of study would be the differential characteristics of *S*s who score consistently high on anxiety scales such as TA, GA, and LP as compared with *S*s who do not display a consistent pattern.

Two sets of correlations of particular interest concern those of the Defensiveness and Hostility scales. With respect to the latter, one is struck by its positive correlation with the three scales purportedly tapping anxiety (GA, TA, LP). If one assumes nonoverlapping definitions of the characteristics measured, the obtained correlations would suggest the possibility that individuals who respond in a hostile manner in interpersonal situations are often reacting to anxiety with which they cannot adaptively cope. Also, of course, in many individuals, the expression of hostility leads to anxiety. The correlation of .56 between the H and LP scales suggest that insecurity and lack of protection in childhood may be an important factor in determining subsequent hostile behavior.

The other striking aspect of the correlation in Table 1 is that concerning the high negative *r*'s between scores on the D scale and all but one of the other scales. This finding lends further evidence to the hypothesis that the tendency to admit to undesirable personal characteristics (e.g., anxiety or hostility) is related to the degree of defensiveness exhibited by the individual. From a different point of emphasis, the anxious individual seems prone to admit to socially undesirable characteristics.

Verbal Conditioning

In order to test the hypotheses concerning the relationship between Autobiographical Survey scores and verbal conditioning, an analysis of covariance was performed for each scale of the survey, using the number of responses of the reinforced class (Verbal Activity) in Trials 1–20 as the control variable and the number of responses of the reinforced class during Trials 21–160 as the criterion variable. For the purpose of these analyses, the distribution of 60 scores for each scale was divided as nearly as possible into lower, middle, and upper thirds.

It had been predicted that *S*s with high scores on the GA, TA, and LP scales would give a significantly greater number of conditioned responses on Trials 21–160 than would *S*s with low scores on these scales. When the three LP groups were compared by analysis of covariance, an $F = 5.13$ ($p < .01$) was obtained. The adjusted means for these groups are presented in Table 2, and examination of them indicates that *S*s with low scores on the LP scale perform at a significantly lower level than do other *S*s in the LP score distribution. Virtually the identical finding was obtained for the TA scale. An $F = 3.49$ ($p < .05$) was obtained in this case; and as Table 2 indicates, the adjusted means for the middle and upper TA groups are significantly higher than for the lower TA group. The results with respect to the upper, middle, and lower GA groups failed to attain significance ($F = 1.68$, $p < .20$). However, the pattern of adjusted means was essentially the same as that obtained for the TA and LP scales. In general, then, the results with respect to these three scales provide support for the hypothesis that *S*s with relatively higher scores tend to adopt responses reinforced by *E* more readily than do low scores on three scales of the Survey.

The prediction had also been made that high scores on the D and H scales would be related to lowered performance levels in the verbal conditioning situation. This prediction was supported in the case of the D scale ($F = 4.65$, $p < .025$). The adjusted means (Table 2) for *S*s in the lower and middle thirds of the distribution of D scores were significantly higher than the mean for the high D group. It thus appears that degree of defensiveness is related to the extent to which *S*s adopt responses reinforced by *E*.

TABLE 2
MEANS ADJUSTED BY COVARIANCE ANALYSIS FOR HIGH, MIDDLE, AND LOW TA, LP, AND D GROUPS

	Adjusted Means		
	High	Middle	Low
TA scale	61.53	60.02	47.59
LP scale	64.90	59.93	47.08
D scale	48.09	59.51	64.14

The results with respect to the H scale, however, were not significant, and the prediction that high H scores would be related to poor verbal conditioning was not supported.

Psychotherapists' Ratings

An exploratory effort was made to determine if, as had been hypothesized, *S*s behave to some extent similarly in the psychotherapy as in the verbal conditioning situation. The method involved having therapists rate their patients on a number of variables, correlating these ratings with Autobiographical Survey scores, and also relating the therapist ratings to verbal conditioning scores.

To the extent that compliance with authority is a means of reducing tension and anxiety, it was predicted that therapists' ratings of the degree of compliance with which a patient accepted interpretations of the therapist would be positively correlated with the anxiety scales of the Survey. If this were so, it would also be expected that *S*s rated as being highly compliant would condition faster than would *S*s rated as being noncompliant. Biserial correlations were obtained between the ratings and scores on GA, TA, and LP for 38 of the 60 *S*s in this experiment. Significance was obtained only for the relationship involving LP ($r_b = .38$), although the GA and TA correlations were also positive and approached significance. Concerning the relation of verbal conditioning scores and therapists' ratings, an analysis of covariance for groups divided into high, middle, and low compliance ratings showed that the low compliance group performed at a significantly lower level in verbal conditioning than did the high and middle groups.

Psychotherapist ratings on two other variables (dependence and hostility as exhibited in psychotherapy sessions) were found to be unrelated to the Survey and also unrelated to performance in the verbal conditioning situation.

DISCUSSION

In reviewing the results of this experiment, cautious optimism would seem to be appropriate. The word optimism is used because the positive results of this study clearly demonstrate the relevance and importance of controlling individual personality differences in verbal conditioning experiments. The present results, together with those reported by Tatz (10), indicate that in learning situations involving reinforcement by *E*, *S*s bring into the experimental situation attitudes and anticipations that directly affect the extent to which they "accept" *E*'s reinforcement or use these reinforcements as a basis for modifying their own behavior.

Further, the majority of the predictions made with reference to the relationship of scores on the various scales of the Survey and scores on the conditioning series were borne out. These predicted relationships are generally assumed clinically and form a basis for such clinical predictions as of therapeutic success. Thus, the present results also lend support to the analogy drawn between *S*'s response to *E*'s reinforcement and a patient's response to a therapist's interpretations. Although the present findings are by no means conclusive, it does seem that a line of attack within the area of research in psychotherapy might be opened which could conceivably make good use of such an analogy. With the many methodological problems besetting the investigator interested in the psychotherapeutic process per se, it may be tactically prudent to concentrate a degree of energy on the study of nonclinical situations which possess characteristics similar to those obtaining in the clinic.

Several relevant questions are raised by the interrelationships found among the various scales of the Autobiographical Survey used in the study. 1. Are the highly related scores based on unclear or overlapping definitions of these personality characteristics? 2. Do these characteristics, if they are adequately defined, lack independence or are certain meaningful relationships present? 3. Might a combination of both 1 and 2 occur? 4. What characteristics differentiate individuals whose scores on the various Survey scales do not conform to the typical interrelationships shown by the groups of *S*s as a whole? For example, what are the characteristics of individuals whose scores on the various types of anxiety scales differ markedly, e.g., high scores on TA but low on GA or LP?

Caution is to be urged in generalizing on the basis of the present findings. It would seem of considerable importance to replicate the present findings with clinical and nonclinical groups. No attempt was made in the present experiment to compare the verbal conditioning performance of groups of patients differing in diagnostic classification. Would, for example, the presumed greater social awareness of neurotics as compared with psychotics reflect itself in a greater degree of verbal conditioning? With respect to a comparison of the over-all performance of the present group of psychiatric patients with a nonpsychiatric population, the present results are comparable to those reported by Sarason (8), who devised the verbal conditioning procedure used in the present study. Studying VA nonpsychiatric patients, she obtained a considerably higher over-all level of conditioning than was obtained in the present study. Quite likely, the differences in patient population between her study and the present one is of great importance in bringing about this difference in level of performance. What precisely are the factors bringing about the differences in level of conditioning in psychiatric and medical patients is at this point unclear. Another possibly relevant factor accounting for this difference between the two studies is the sex difference between experimenters.

SUMMARY

1. The following measures were available for a group of psychiatric patients in active treatment: (*a*) Scores on the six scales (General Anxiety, Test Anxiety, Lack of Protection, Defensiveness, Need for Achievement, and Hostility) of a personality inventory, the Autobiographical Survey. (*b*) Ratings by the patients' psychotherapists of the degree of compliance, dependence, and hostility displayed by the patients in therapy sessions. (*c*) Level of performance attained in a verbal conditioning task in which *E* systematically reinforced all responses made by *S* which fell in the category of Verbal Activity (i.e., verbs indicating use of the vocal apparatus).

2. The six scales of the Autobiographical Survey were found to possess reasonable reliabilities. Intercorrelations among the six scales indicated positive relationships among all the scales except the Defensiveness scale, which correlated negatively wth the other scales.

3. High scores on the Test Anxiety and Lack of Protection scales were found to be significantly associated with higher levels of verbal conditioning. On the other hand, high Defensiveness scores were found to be related to relatively poor verbal conditioning.

4. Patients rated by their psychotherapists as being very compliant were found to perform in the verbal conditioning situation at a significantly higher level than patients rated as being relatively noncompliant.

5. The results were discussed in terms of previously made predictions and in terms of the feasibility of studying factors affecting behavior in psychotherapy in simpler, more easily controllable situations.

REFERENCES

1. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.
2. FREUD, S. *The problem of anxiety.* New York: W. W. Norton, 1936.
3. GORDON, E. M., & SARASON, S. B. The relationship between "test anxiety" and "other anxieties." *J. Pers.*, 1955, **23**, 317–323.
4. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
5. HATHAWAY, S. R., & MCKINLEY, J. C. *Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1943.
6. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
7. MARK, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 185–189.
8. SARASON, B. R. The effects of verbally conditioned response classes on post-conditioning tasks. Unpublished doctoral thesis, Indiana Univer., 1956.
9. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.
10. TATZ, S. Symbolic mediation in "learning without awareness." Unpublished doctoral thesis, Yale Univer., 1956.
11. TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
12. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1944.
13. VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.
14. VERPLANCK, W. S. The operant, from rat to man: An introduction to some recent experiments on human behavior. *Trans. N. Y. Acad. Sci.*, 1955, **17**, 594–601.

Received May 31, 1957.

INTERRELATIONS AMONG MEASURES OF PERSONALITY ADJUSTMENT IN NONCLINICAL POPULATIONS[1]

FRED E. FIEDLER, JOAN S. DODGE,[2] ROBERT E. JONES,[3] AND EDWIN B. HUTCHINS

University of Illinois

A LARGE number of studies, reported in the literature, require the evaluation of psychological adjustment or of changes in adjustment, whether the study deals with consequences of situational factors, or of psychotherapeutic or psychosurgical intervention.

Most theoretical and empirical work in this area treats the person's adjustment as falling on a unidimensional scale. Thus, Sappenfield writes, "Adjustment involves a continuous variable, and therefore, the evaluation of individuals in terms of this variable should not be limited to the two extremes, [of adjusted and maladjusted]" (19, p. 24). A similar point of view is expressed or implied by other writers in this area (e.g., 2, p. 15; 6, p. 554; 23, p. 102).[4] It has gained considerable theoretical support from the writings of Goldstein (10) and Rogers (17, p. 510) who see the individual reacting to his environment as a total organism, and who see poor adjustment in one area of life as necessarily affecting all other areas of personality. Improvement or deterioration of adjustment should, therefore, take place on all fronts.

Whether adjustment is in fact unidimensional in character, or whether a person may achieve adjustment in one area of his life irrespective, or at the expense of, his adjustment in another is a question for research to which the present report hopes to contribute. It asks how various adjustment indices intercorrelate with one another in samples of subjects (*S*s) whose adjustments, as measured by a variety of indices, range from excellent to relatively poor.

Other workers in this area have previously raised this question. Ellis (5) has shown that objective personality questionnaires are of dubious value in distinguishing between adjusted and maladjusted individuals, and that intercorrelations among similar personality questionnaires are generally low or insignificant. Tindall (21), in a more recent paper, likewise reports low intercorrelations among 16 adjustment indices. On the other hand, Tucker (22) reports positive correlations among change in adjustment measures, and Cartwright and Roth (1), working with post-therapy ratings and indices, have found that 18 of 45 correlations among adjustment scores were greater than .46 (beyond the .01 level) although pretherapy scores do not intercorrelate. To summarize, the status of our knowledge concerning indices of adjustment is by no means clear. The present report deals with a fairly wide variety of frequently used measures of adjustment and change in adjustment, and it compares results from four studies which used subjects (*S*s) who were unselected in terms of clinical symptomatology.

The data were gathered in connection with a research program on "quasi-therapeutic" relations in small groups. The studies were conducted on the assumption that any interpersonal relationship might be helpful or detrimental to adjustment, and that the adjustive relations could be differentiated from those which would lead to deterioration of adjustment.

The specific design of the various studies will be considered in a forthcoming report. The present paper limits itself to a brief description of the samples, the methods, and the research setting in each of the four studies; and the presentation and interpretation of data relevant to our inquiry.

METHOD

The Four Studies

We studied four different groups, two of which consisted of college students, and two of military personnel, viz., army tank crews and anti-aircraft artillery (AAA) units. Data were collected twice in each of the studies, so that we could obtain scores on adjustment indices at the first and second testing sessions, as well as scores indicating changes in adjustment.

---

[1] This study was conducted by the Group Effectiveness Research Laboratory, Department of Psychology, University of Illinois, under Contract DA-49-007-MD-569 with the Office of the Surgeon General. Reproduction in whole or in part is permitted for any purposes of the United States government.

[2] Now at the Graduate School of Business and Public Administration, Cornell University.

[3] Now at the VA Mental Hygiene Clinic, Huntington, W. Va.

[4] This is, of course, also the underlying assumption wherever a number of adjustment scores are averaged or combined.

The Dormitory study. The first study was conducted on a group of 107 male college students living in a new university dormitory.[5] Most *S*s had entered college at the beginning of the semester in which the data were collected. About half of the students were freshmen, the remainder being transfer and incoming graduate students. *S*s were assigned to their rooms in order of their application for housing. All *S*s were tested during the fifth week of the semester, and again during the eleventh week. Our analyses are based on 87 *S*s for whom complete data were available.

The Barracks study. Three years after the Dormitory study, similar data were collected on a sample of students living in 13 temporary barracks. Each barracks, which is divided into four-man suites, holds a maximum of 16 men. Although a total of 172 *S*s were tested, analyses were restricted to the 102 men for whom complete data could be obtained at both testing sessions. These *S*s were tested during the third and twelfth weeks of the semester.

The Tank Crew study. This investigation was conducted on members of a tank battalion during the crew training phase. *S*s were enlisted men with an average age of 22 years. Of the 160 men initially tested, 49 were married, three divorced. Average education was 9.7 years, mean Army General Classification Test score was 90. All men were new to each other and to the particular type of tank training. This group was thus considerably more heterogeneous than the student groups in the previously mentioned studies. Upon retest, the sample size was reduced from 160 to 95 due to transfers, discharges, and other causes.

The Anti-aircraft Artillery study. The last group of *S*s came from operational AAA units. Here again there was considerable heterogeneity among *S*s. Average age of the men, all of whom were enlisted, was 23 years, average education 11.2 years, mean Army General Classification Test score was 101. Of the 200 men in our sample, 36 were married.

The AAA crews were "old" units, that is, these men were no longer in training, but were performing their regular duties within the Air Defense Command. As with tank crews, the loss of data, due to furloughs, sick call, or turnover in our particular sample was considerable. In the 12-week period between testings, the original sample shrank from 485 men to 200 for whom complete data were available. Our analyses are based on this group of 200 *S*s.

Criteria of Adjustment

As a rule, the development of adequate criteria of adjustment presents a very difficult problem. One major source of this difficulty arises from the fact that each theoretical orientation has defined adjustment or integration on the basis of different assumptions and by different operations. However, it is perhaps somewhat easier to develop adequate adjustment measures in a group setting than when the individual is considered by himself, as is the case in most studies on individual therapy. The individual in a group situation is almost continuously interacting with others, and he is thus under almost constant observation by his fellow group members. In addition, each man shares the environmental stresses of others in the group, and his reactions to these stresses are therefore comparable to those of other group members.

We shall here deal with three classes of criteria. These are (*a*) subjective indices of adjustment which depend on *S*'s self-reports, (*b*) group acceptance criteria which are based on group members' evaluations or descriptions of *S*, and (*c*) personal effectiveness criteria indicative of the individual's ability to operate effectively. Most of these indices, or similar ones, have been used by other investigators in this area.

Subjective indices of adjustment. These criterion scores were derived from *S*'s self descriptions. In particular we employed scale sheets on which *S* described himself as he thought he really was, and how he would ideally like to be (his ideal self), and he described various other members of his unit. These descriptions were made on a series of identical scale sheets, each of which contained the same 20 items.[6] Items were six-point scales bounded by pairs of polar adjectives. For example:

Friendly ___:___:___:___:___:___ Unfriendly
Stable ___:___:___:___:___:___ Unstable

In the Dormitory and Tank Crew studies, *S*s described all other men in their units; in the Barracks and AAA studies, where each unit was considerably larger, *S*s were asked to describe only a limited number of men. As we shall discuss below, these descriptions of others were used as indices of group acceptance.

Self-esteem scores could be obtained from self-descriptions by determining the average favorability of *S*'s self rating. As will be noted, each of the 20 items can easily be polarized: it is better to be friendly than unfriendly, stable than unstable, and so forth. A person who consistently describes himself in a more favorable manner is presumed to have higher *self-esteem*. Obviously, as is the case with all self reports and inventories, this score may reflect very superficial attitudes, such as a facade, or it may reflect a genuine degree of self-esteem on the part of the *S*.

Using the measure D (4, 15) we obtained a second index, called *Self Satisfaction*, which is defined as the degree to which *S*'s self-description is similar to his description of how he ideally would like to be. This index has been utilized by one of the authors (Fiedler, 7) in studies of patient-therapist interactions. It has been extensively used by Rogers and his co-workers to demonstrate movement in client-centered therapy (1, 16).

In addition, an abbreviated form of the Taylor Manifest Anxiety Scale (20), which has become one of the most popular inventories in recent studies of adjustment (9, 12), was given to all *S*s save those in the Dormitory sample. In the AAA study, the *General Army Adjustment* scale, described by Havron, Fay, and Goodacre (11) was also administered.

Group acceptance criteria. Two measures were obtained. One, *Esteem by Others*, is derived from rating scales on which *S*s were asked to describe others in their group. It is analogous to the Self-Esteem score, and again is the average item score with which a particular *S* is described. The second index, *Sociometric Status*, was obtained by asking *S*s to name those in their group whom they preferred most, in whom they would confide personal problems, whom they would choose as a leader, etc. A pooling of these highly intercorrelated items yielded a sociometric preference index. The more an *S* was chosen, the higher is his status in the group. (Cf. French (8), Pepinsky, Siegel, and Van Atta (16), and Cox (3) whose studies have related sociometric status to some aspects of adjustment.)

---

[5] We are indebted to Melvin Manis for permission to utilize his data for purposes of our project. His own study is reported in (14).

[6] The Dormitory study used seven-point, 24-item scales.

Personal effectiveness criteria. Three personal effectiveness criteria were available for the college student samples. The first of these is based on the frequency of the individual's *Health Center* visits. It assumes that the anxious, maladjusted person will experience relatively more symptoms of a psycho-physiological nature than will one who is adjusting. He will also be more aware of his aches and pains, partly because of hypochondriacal defense tendencies, partly because general anxieties will tend to augment pain through increased tension.

Student Counseling Bureau visits for personal maladjustments were similarly tabulated for our *S*s as an indication that the individual feels that he cannot solve his problems without professional help.

Finally, we collected students' *Grade Point Averages* (GPA) for the semester of our study. In the Barracks study, we were also able to obtain mid-semester grades which are given to freshmen on an advisory basis. Collection of these data permitted us to compute change in grade point average for this sample.

In the AAA units, adequate individual proficiency scores were not available. Sick-call data were obtained but, for a variety of reasons, such as unequal accessibility of the dispensary to various units, they do not appear to be adequate for research purposes. Ratings on disciplinary offenses appeared to be usable, however. These data were obtained in the AAA study and provide some indication of the individual's ability to adjust to the Army.

Measures of Change in Adjustment

As indicated above, all *S*s were tested twice, with intervals between testing sessions ranging from six weeks to 13 weeks in the various samples. Since any change score necessitates two measures on the same individual, these scores are available only for subjective indices, group acceptance criteria, and for grade point average for the Barracks study.

Reliability of Scores

The relation between two indices of adjustment cannot be interpreted adequately unless we know the reliability of each of the scores. Table 1 lists the corrected split-half reliabilities. Only one estimate is here presented. Where more than one reliability was computed, the estimate based on the larger *N* is shown. Relatively little variation was found from sample to sample. Since some scores, such as Health Center Visits or Counseling Bureau Visits, do not lend themselves to this type of statistical analysis, these indices must be interpreted with considerably more caution.

TABLE 1

SPLIT-HALF RELIABILITY COEFFICIENTS[a] OF PERSONALITY ADJUSTMENT INDICES

Index	Sample	*N*	*r*
Self-satisfaction	Barracks	169	.76
Self-esteem	AAA	397	.83
Mean Esteem by Others	Dorms	86	.70
Sociometric status	Dorms	99	.94
Taylor Anxiety	AAA	200	.91
General Army Adjustment	AAA	200	.73

[a] Coefficients are the corrected *r*'s based on the Spearman Brown Prophecy formula.

CORRELATIONS AMONG CRITERIA OF LEVEL OF ADJUSTMENT

If adjustment is to be treated as a unitary trait or general attribute, it is expected that specific measures of its various operational definitions will be correlated. An individual showing superior adjustment on one index should be above average on others as well. A lack of such co-variance would suggest that various adjustment indices measure very specific components of adjustment, or else, that adjustment in one area of behavior does not necessarily imply adjustment in a different area.

Intercorrelations of Adjustment Indices at Time II

Table 2 presents the intercorrelations among the various adjustment indices at the time of the second testing. As can be seen, all correlations are very low, even when highly reliable internal criteria, such as Taylor Anxiety scores and self-satisfaction or self-esteem scores, are involved.

It must be recognized, of course, that some maladjustment symptoms are substitutive rather than supplemental or additive. A person who goes to the Health Center for psychosomatic symptoms is less likely to consult the Counseling Bureau as well. Yet, we would expect that someone who seeks help at the Counseling Bureau or the Health Center or who performs poorly in school would be dissatisfied with himself. As can be seen, however, this latter expectation is not borne out by our data.

Although relationships between indices presented here are not artifactual in the mathematical sense, all indices based on self-description are psychologically related. A person who

TABLE 2

SUMMARY OF RELATIONSHIPS BETWEEN INDICES OF ADJUSTMENT OBTAINED AT THE SECOND TESTING

Indices Related	Dorms $N = 87$	Barracks[d] $N = 102$ [$N = 71$ (Est. by O.)]	Tanks $N = 95$ [$N = 52$ (Est. by O.)]	AAA $N = 200$
Self-Satisfaction and:				
Self Esteem	.60[a]**	.55[a]**	.36[a]**	.44[a]**
Sociometric Status	.09	.09	.00	.10
Mean Esteem by Others	−.02	.15	−.14	−.09
Taylor Anxiety	—	.34[b]*	.37[b]**	.30[b]**
Health Center Visits	.00	.19	.10	—
Grade Point Average	.17	.30	—	—
Counseling Bureau Visits	.15	.16	—	—
Annoyance Scale	—	—	.08	—
Disciplinary Ratings	—	—	—	.01
General Army Adjustment	—	—	—	.15*
Self Esteem and:				
Sociometric Status	.20	.10	.04	.09
Mean Esteem by Others	.22*	.09	−.20	.02
Taylor Anxiety	—	.25[b]	.43[b]**	.34[b]**
Health Center Visits	.22*	−.06	−.09	—
Grade Point Average	−.14	.28	—	—
Counseling Bureau Visits	.08	.22	—	—
Annoyance Scale	—	—	−.10	—
Disciplinary Ratings	—	—	—	.11
General Army Adjustment	—	.	—	.16*
Sociometric Status and:				
Mean Esteem by Others	.38[c]**	.42[c]*	.34[c]*	−.17[c]*
Taylor Anxiety	—	−.09	.09	.04
Health Center Visits	.06	−.04	.02	—
Grade Point Average	.05	.31*	—	—
Counseling Bureau Visits	.10	.00	—	—
Annoyance Scale	—	—	.01	—
Disciplinary Ratings	—	—	—	.33**
General Army Adjustment	—	—	—	−.03
Mean Esteem by Others and:				
Taylor Anxiety	—	−.02	−.13	−.07
Health Center Visits	.15	.08	−.09	—
Grade Point Average	.03	.14	—	—
Counseling Bureau Visits	.23*	.06	—	—
Annoyance Scale	—	—	−.07	—
Disciplinary Ratings	—	—	—	.24**
General Army Adjustment	—	—	—	.12
Taylor Anxiety and:				
Health Center Visits	—	.16	.07	—
Grade Point Average	—	.18	—	—
Counseling Bureau Visits	—	.05	—	—
Annoyance Scale	—	—	−.01	—
Disciplinary Ratings	—	—	—	.04
General Army Adjustment	—	—	—	.27**
Health Center Visits and:				
Grade Point Average	−.10	−.06	—	—
Counseling Bureau Visits	.06	.02	—	—
Annoyance Scale	—	—	−.07	—
Grade Point Average and:				
Counseling Bureau Visits	.03	.27	—	—
Disciplinary Ratings and:				
General Army Adjustment	—	—	—	.11

[a] Spuriously high relation based on self-descriptions.

[b] Both variables are self-descriptions.

[c] Both variables based on descriptions by the same others.

[d] r's estimated from ϕ coefficients.

* Significant at the .05 level.

** Significant at the .01 level.

describes himself as well adjusted on the Taylor Manifest Anxiety Scale is likely to describe himself also as favorable on the self-description sheet. Hence, his Self-Esteem score, and his Self-Satisfaction, will also be relatively high. Under these conditions, one would actually expect higher correlations than occurred here. Similarly, Sociometric Status and Esteem by Others are psychologically related both being ratings given to *S* by his fellow group members. Since they are based on responses from the same raters, it is again noteworthy that the relations are so low. Inspection of scatter diagrams reveals no evidence of curvilinearity in our data.

Intercorrelations were also computed for indices obtained at the first testing session. These are essentially similar to those found in Table 2. (A comparison is presented in Table 4 which summarizes the median intercorrelations obtained at the first and second testing session and for change scores which are presented below.)

As can be seen from Table 2, there is relatively little similarity in the pattern of intercorrelations from study to study. These data,

TABLE 3

SUMMARY OF RELATIONSHIPS BETWEEN INDICES OF ADJUSTMENT (CHANGE SCORES)

Indices Related	Dorms $N = 87$	Barracks^{d} $N = 66$	Tanks $N = 95$ [$N = 52$ (Est. by O.)]	AAA $N = 200$
Self Satisfaction and:				
Self Esteem	.46^{a**}	.43^{a**}	.18^{a}	.38^{a**}
Sociometric Status	−.11	−.10	.00	.08
Mean Esteem by Others	.15	.03	.09	.04
Taylor Anxiety	—	.26^{b*}	.24^{b*}	.08^{b}
Health Center Visits	—	—	.18	—
Annoyance Scale	—	—	.10	—
General Army Adjustment	—	—	—	.01
Grade Point Average	—	−.12	—	—
Self Esteem and:				
Sociometric Status	.08	−.14	.01	.04
Mean Esteem by Others	.00	.23	−.13	−.02
Taylor Anxiety	—	.23^{b}	.01^{b}	.09^{b}
Health Center Visits	—	—	−.09	—
Annoyance Scale	—	—	−.10	—
General Army Adjustment	—	—	—	.00
Grade Point Average	—	−.08	—	—
Sociometric Status and:				
Mean Esteem by Others	.24^{c*}	.09^{c}	.16^{c}	−.09^{c}
Taylor Anxiety	—	.26*	.05	−.03
Health Center Visits	—	—	−.15	—
Annoyance Scale	—	—	.10	—
General Army Adjustment	—	—	—	.00
Grade Point Average	—	.19	—	—
Mean Esteem by Others and:				
Taylor Anxiety	—	.04	.11	−.03
Health Center Visits	—	—	−.11	—
Annoyance Scale	—	—	.04	—
General Army Adjustment	—	—	—	.01
Grade Point Average	—	.01	—	—
Taylor Anxiety and:				
Health Center Visits	—	—	.13	—
Annoyance Scale	—	—	.13	—
General Army Adjustment	—	—	—	.25**
Grade Point Average	—	.27*	—	—
Health Center Visits and:				
Annoyance Scale	—	—	.01	—

[a] Spuriously high relation based on self descriptions.
[b] Both variables are self descriptions.
[c] Both variables based on descriptions by the same others.
[d] r's estimated from ϕ coefficients.
* Significant at the .05 level.
** Significant at the .01 level.

TABLE 4
SUMMARY OF MEDIAN CORRELATIONS[a] OBTAINED FROM THE FIRST AND SECOND TESTING SESSIONS AND FROM CHANGE SCORES

	Dorms	Barracks	Tanks	AAA
First testing session	.16	−.05	.07	.11
Second testing session	.10	.14	.01	.11
Change scores	.08	.07	.05	.01

[a] In determining the median correlations *r*'s which were spuriously high were not included.

as well as the generally low intercorrelations, indicate that factor analyses would not contribute materially to their interpretation.

Interrelations among Change in Adjustment Scores

Although we have not found relationships among level of adjustment scores, it is nevertheless possible that changes in adjustment over various aspects of personality might be related. Improvement on one criterion score could be accompanied by improvement on others irrespective of the level at which *S*s are observed to be at any one time. To determine whether or not co-variations in change scores do exist, all available change scores were intercorrelated. Table 3 presents these findings.[7]

If we exclude the relations which are artifactual in nature, the correlations are again uniformly low. In our samples, then, there is no evidence that improvement (or deterioration) on one criterion is accompanied by significant changes in other types of criterion scores. Nor are the relations very similar if we compare the profile or correlations in one study with that in another.

Obviously, the criteria of change do not vary together. A person who improves in his subjective feelings may or may not act in a more adjusted manner as judged by others, and he may or may not perform more adequately or efficiently in performing his job. A few individuals undoubtedly do change in all aspects of adjustment, but the possible identification of these individuals is a question for further research.

DISCUSSION

While there is, of course, no logical reason to expect correlation between measures such as sociometric status and self satisfaction, both are frequently considered indices of adjustment. Table 4 summarizes the median intercorrelations among the criterion scores in these studies. Less than half are above .10 and none exceed .16. Text books do not generally discuss the possibility that Adjustment A may be totally different from Adjustment B. Although Rogers and Dymond (18, p. 431–432) explicitly recognize this possibility (which contrasts with Rogers' earlier view), we cannot readily find this viewpoint in standard psychiatric texts, or in many of the texts on personality theory or abnormal psychology.

It is especially noteworthy that Cartwright and Roth (1) have reported fairly high posttherapy intercorrelations in clinically selected *S*s but low relations in pretherapy scores. This observation may point to the possibility that the structure of posttherapy adjustment in individuals who complete counseling or therapy may differ from that of *S*s who never felt the need for treatment. It could be possible that the individuals in therapy experience a more basic kind of anxiety, interfering more comprehensively with life adjustment problems. It is also possible that therapy typically aims for, and achieves, a rebalancing of defenses so that the individual is similarly adjusted in all areas of living. In any case, these discrepancies need to be further examined.

The generalizability of these findings may be challenged on the grounds that the sample of criteria chosen was not adequately representative of all adjustment areas. This possible criticism is, of course, a matter of definition. We recognize that we did not include many excellent adjustment inventories or indices which have been utilized in the past. However, the fact that Tindall's (21) and Ellis' (5) findings agree with ours adds to the weight of our evidence, as does the fact that our findings are very similar in four different samples of *S*s.

A more important question is whether the time span between the first and the second testing is sufficiently long enough to permit us to see co-variations in change scores. We recognize this to be a definite limitation of our data, although studies not yet reported have shown that some "quasi-therapeutic" changes do take place within this time.

---

[7] It should be pointed out that change scores are subject to "regression to the mean" effects which can seriously distort our estimates of the "true" gain scores. Lord (13) has pointed out that these effects will be considerable when the standard deviations differ greatly at Time I and at Time II. Since this is not the case in our studies, it does not affect materially the correlations presented in this paper.

SUMMARY

This paper has presented intercorrelations among a variety of indices which have been used as measures of personality adjustment. The most important fact that emerges is the general lack of correlation among different indices—even among those which are reliably measurable and which could be expected to correlate with each other. Thus, our data yield no evidence justifying the assumption that adjustment, in its present state of definition, should be considered a unitary trait in clinically unselected populations.

In light of our data, we must recommend that subsequent research be confined to specific measures when inferences are to be drawn about personality adjustment or personality changes. It should also be recognized that results from clinically selected, or post-therapy groups, may be different from those obtained on *S*s who are unselected from the point of view of presenting symptoms, or the fact that they expressed a need for psychotherapeutic help.

REFERENCES

1. CARTWRIGHT, D. S., & ROTH, I. Success and satisfaction in psychotherapy. *J. clin. Psychol.*, 1957, **13**, 20–26.
2. COLEMAN, J. C. *Abnormal psychology and modern life.* (2nd ed.) New York: Scott Foresman, 1956.
3. COX, F. N. Sociometric status and individual adjustment before and after play therapy. *J. abnorm. soc. Psychol.*, 1953, **48**, 354–356.
4. CRONBACH, L. J., & GLESER, GOLDINE C. Similarity between persons and related problems of profile analysis. *Psychol. Bull.*, 1953, **50**, 456–473.
5. ELLIS, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, **42**, 385–440.
6. FENICHEL, O. *The psychoanalytic theory of neuroses.* New York: Norton, 1945.
7. FIEDLER, F. E., & SENIOR, KATE. An exploratory study of unconscious feeling reactions in fifteen patient-therapist pairs. *J. abnorm. soc. Psychol.*, 1952, **47**, 790–796.
8. FRENCH, R. L. Sociometric status and individual adjustment among naval recruits. In D. Cartwright & A. Zander (Eds.), *Group dynamics.* Evanston, Ill.: Row Peterson, 1954.
9. GALLAGHER, J. J. Manifest anxiety changes concomitant with client centered therapy. *J. consult. Psychol.*, 1953, **17**, 443–446.
10. GOLDSTEIN, K. *Human nature in the light of psychopathology.* Cambridge: Harvard Univer. Press, 1940.
11. HAVRON, M. D., FAY, R. J., & GOODACRE, D. M., 3rd. *Research on the effectiveness of small military units.* PRS Report No. 885. Institute for Research on Human Relations, 1951.
12. HOLTZMAN, W. H., CALVIN, A. D., & BITTERMAN, M. E. New evidence for the validity of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853–854.
13. LORD, F. M. The measurement of growth. *Educ. psychol. Measmt*, 1956, **16**, 421–437.
14. MANIS, M. Social interaction and the self concept. *J. abnorm. soc. Psychol.*, 1955, **51**, 362–370.
15. OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.
16. PEPINSKY, H. B., SIEGEL, L., & VAN ATTA, E. L. The criterion in counseling: A group participation scale. *J. abnorm. soc. Psychol.*, 1952, **47**, (Suppl.) 415–419.
17. ROGERS, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
18. ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.
19. SAPPENFIELD, B. *Personality dynamics.* New York: Knopf, 1956.
20. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
21. TINDALL, R. H. Relationships among indices of adjustment status. *Educ. psychol. Measmt*, 1955, **15**, 152–162.
22. TUCKER, J. E. Measuring progress in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* Pennsylvania State Coll., 1953.
23. WHITE, R. W. *The abnormal personality.* New York: Ronald Press, 1956.

Received May 31, 1957.

INTERACTING EFFECTS OF CONTROL, MOTIVATION, GROUP PRACTICE, AND PROBLEM DIFFICULTY ON ATTEMPTED LEADERSHIP[1]

BERNARD M. BASS, MARGARET W. PRYER, EUGENE L. GAIER, AND AUSTIN W. FLINT
Louisiana State University

THE differentiation of *attempted* from successful and effective leadership was suggested by Hemphill[2] and applied in the construction of two distinct theories of leadership, one by Hemphill and one by Bass (2). For the latter, leadership is attempted whenever one member of a group behaves in a manner directed toward changing any behavior of another. Hemphill's definition is more restricted. Leadership attempts refer only to acts by one member aimed at restructuring the existing pattern of interaction between that member and others.

The broader conceptualization permits the acceptance of almost any response in initially leaderless discussions as attempted leadership. Participation in leaderless discussions is equated with attempted leadership, since it is assumed that almost all talk in initially leaderless discussions is directed toward influencing other members primarily by giving opinions, suggestions, information, and by summarizing, clarifying as well as organizing, the topic discussed (1).

One member, A, *controls* another, B, if he can give or get for B what B wants. If B's wants are strong, A has *power* to change B's behavior by giving or denying B what B wants. Therefore, A can successfully lead B, if leadership behavior is defined as acts directed toward changing another's behavior. But A will be successful only if B's motivation to obtain what A can give is fairly high.

If A expects to be successful, he is more likely to attempt leadership. Hence, under conditions of high motivation, A's control results in power to lead B and A will attempt to do so. However, if motivation of A and B are low, A's control will not yield power and A will not attempt more leadership.

A second cause-effect mechanism that relates control and motivation produces the same outcome: Member A may be expected to use his high control only if he is highly motivated. Either rationale supports the deduction that under conditions of strong motivation, high control members will attempt more leadership than low control members. Under conditions of weak motivation, differences should become negligible or completely disappear.

The major purpose of the present study was to test this hypothesis that control yields attempted leadership but only under conditions of high motivation. A second purpose was to examine the effects of group practice on the relations of control, power, and attempted leadership. Continued experience was expected to reinforce the attempts of high power members to lead.

A third purpose was to examine the effects of problem difficulty on attempted leadership. Easy problems require relatively little leadership, and little should be attempted; but attempted leadership should be more frequent, especially by the more powerful members, as problems become more difficult (but not so difficult that the group withdraws from trying to solve them).

METHOD

Manipulating Motivation by Selection

On an a priori basis, it was believed motivation could be varied most successfully by classification of *S*s according to their varying motivation to attain some important life goal. At the end of the sophomore year, the land-grant college student has fulfilled the requirements of the federal government for participation in ROTC training. However, if the student chooses, and is found qualified, he may remain in the ROTC program two additional years, and receive an officer's commission at the end of this period. Students vary widely in their desire to enter Advanced ROTC. We obtained varying levels of motivation for test performance by subjecting students differing in stated desire to enter Advanced ROTC, to an ROTC "screening test." To sample ade-

---

[1] This study was supported by Contract N7 ONR 35609, Group Psychology Branch, Office of Naval Research.

[2] Unpublished manuscript by J. K. Hemphill, "A proposed theory of leadership in small groups."

quately students with low motivation, all students, regardless of motivation, were ordered to take the test. *S*s were instructed that the "teamwork test" might aid in determining whether or not the student was qualified for acceptance for advanced training. It was expected that students strongly interested in entering Advanced ROTC would be more highly motivated to perform well in the test situation than students who lacked such interest.

About two months before the present study was initiated, the entire sophomore class was given a questionnaire composed of seven five-step rating scales concerning patriotism and military service. Six were "dummy" scales. The key item was: "How important is it for you to enter Advanced ROTC?" The 342 sophomores who responded, did so in the following percentages:

Response	Percentage Responding
A. A great deal	21%
B. Fairly much	18
C. To some extent	34
D. Comparatively little	11
E. Not at all	16

Several analyses support the inference that response to the key question provided a valid assessment of motivation:

Attitude toward military career. Response to the key question was consistent at the .01 level of confidence with response to the item "Right now, how desirable is a career in the military service to you?"

Application for advanced ROTC. The following percentage of each category or response to the key question actually applied for Advanced ROTC during the next three months following completion of the questionnaire:

	Response to Key Question				
	A	B	C	D	E
Percentage applied	80	71	43	17	0
Percentage did not apply	20	29	57	83	100

For experimental use, the number of motivation levels was reduced to three approximately equal samples as follows:

Motivation Level	Responses	Percentage Responding
High	A and B	39%
Medium	C	34
Low	D and E	27

Failure to report. The lower two levels of motivation were less likely to appear for the "screening" examination. In fact, in order to reduce to a minimum the loss of groups, because someone did not appear for testing, it was necessary to schedule two alternate *S*s for the Medium (C) and Low (D and E) motivation level groups, while only one alternate was necessary for the high level (A and B) groups. Despite this precaution, the following percentages of scheduled groups were not tested because of failure to appear of at least five men *including* the alternates.

Motivation Level	Responses	Percentage of Scheduled Groups Lacking Five Men to Be Tested
High	A and B	13.3%
Medium	C	20.0
Low	D and E	46.7

Fortunately, many of the men from the aborted groups could later be rescheduled in different groups.

Reported effort. About a week following the last testing session, a questionnaire was sent to all participating *S*s. Of these, 198 returned the questionnaire. Motivation level was significantly related to effort according to self-ratings to the question "How hard would you say you tried to do as well as possible on the teamwork test?" Thus, 59% of the "highs" reported they had tried a great deal, 42% of the medium sample so reported and only 27% of the "lows" indicated such effort.

Griping behavior. Following a suggestion drawn from Willerman's (6) work on involvement in groups that low motivation *S*s would place emphasis on taking the test, while *S*s with high motivation would emphasize high score barriers, *S*s were asked on the posttest questionnaire to comment on any aspect of the testing procedure. A content analysis found that the "highs" were more likely (at the .01 level of confidence) to complain about difficulties in earning a high score while "lows" griped about having to take the test. "Lows" were also less likely to return the questionnaire.

Manipulating Control

Amount of control of individual members as well as their variance in control were manipulated by experimental treatment of their status.

Amount of control. Before actual testing, each of the five members of the group drew from a deck, a card with one to four "Xs" marked on it. The number of "Xs" a member drew signified how much control he had. All were informed that the number of "Xs" determined how much weight would be carried by each *S*'s evaluation of the other members' performance at the end of the screening examination. For example, a member who drew four "Xs" would carry four times as much weight in determining everyone else's final ratings on adequacy of test performance, as compared with a person who drew one "X."

Variance control. The extent to which members varied in control within a given group was regarded as a significant determiner of how members would react. The greater the differentiation among members, the more distinct should be differences in behavior due to control. To manipulate *control variance*, the decks of cards were "stacked" in one of three ways. In one-third of the groups tested, the five members drew from a 20-card deck where every card was marked with a single "X." In one-third of the groups, the five members drew from a five-card deck which contained three cards with one "X," one card with two "Xs" and one card with three "Xs." In the remaining one-third of the groups, the five members drew from a five-card deck which contained four cards with single "Xs" and one card with four "Xs." The three types of groups had variances in control of 0.00, 0.64, and 1.44 respectively.

Measuring Attempted Leadership

Attempted leadership was defined as the amount of time spent talking as measured by a "throat microphone-to-electric clock" circuit. As soon as the *S*s were seated for testing, each *S* fastened around his neck a throat microphone, which was connected to an electric clock circuit. The time in seconds for each of ten brief problem-solving discussions was recorded by an assistant.

The Group Task

Each group was administered ten problems. In each problem, the five members first privately ordered their opinions of the correct rank order of the familiarity of five words. They then carried on a group discussion to reach a group decision about the ranks. Again, they privately ranked the words. The correct rank order according to a word familiarity list constructed by Noble (5) was then reported to the participants (3).

Problem difficulty. An easy problem was constructed by assembling from Noble's list words that widely differed in familiarity and had a high mean familiarity as a set. A difficult problem was constructed using unfamiliar words differing little from each other in their familiarity. The problems were pretested on 100 Ohio State University freshmen. On the basis of the average correlation between each of these *S*'s initial private rank orderings of the five words of each problem, pairs of problems at each of five levels of difficulty were identified.

Randomized latin squares were used to provide each group with a random sequence of problems of varying difficulty. One additional restriction was imposed. If a problem appeared in the first half of the ten problems presented to a group, its equally difficult mate appeared somewhere in the second half. This provision was to insure that the overall difficulty of the second half of the test would be comparable to the difficulty of the first half.

RESULTS

Original plans called for assessing 90 groups, 30 at each of the three levels of motivation. Practical considerations permitted the testing of only 60 groups, and in order to maintain proportionality to permit an unbiased two-factor analysis of variance of the effects of control and motivation on attempted leadership, we withdrew nine groups randomly (within the restrictions of the design) leaving 51 in the final analysis, distributed as shown in Table 1.

The data analysis will be presented in two parts. First, intergroup differences in attempted leadership (or total amount of participation) under the varying conditions of motivation and control variance will be tested. Then we will consider how an individual differed from his own group mean in attempted leadership as a consequence of his amount of control over the others in his own group. In the first case we will examine the absolute differences between groups of different control variance ignoring relative variations within the groups. In the second type of analysis, we will examine the relative differences between members of different control, ignoring the absolute differences between groups. In both cases, we will examine trends in obtained relations as a function of group practice and problem difficulty.

Analysis of Variance of Groups

Table 2 displays the average member's attempted leadership or amount of participation in the 51 groups under one of the three levels of motivation and three levels of variance in control averaged for all ten problems. According to Type III (4) analyses of variance, attempted leadership varied significantly at the .01 level with motivation. The "lows" in motivation talked only 58 per cent as much as the "highs," the "medium motivated" talked only 74 per cent as much as the "highs." The differences due to variance in control, to problem difficulty and to group experience were accounted for by the fluctuations between groups or within groups on repeated trials.

We expected that decisions would be reached most quickly with least total participation under high variance in control, but the trend—although not significant—was in the reverse direction. High variance in control, particularly under high motivation, yielded the greatest total amount of attempted leadership. The absolute differences were quite large.

Practice effect. We expected an increase in participation under high motivation as members became more familiar with each other, and a decrease in participation from feelings of apathy and boredom among those low in motivation as the test progressed. The trends were suggestive but failed to attain statistical significance. It was surmised that since each individual test problem was unreliable, comparing performance on the first half of the test with performance on the second half of the test might still reveal a statistically reliable trend. The Type III analysis of variance was recomputed to contrast the results of the first five trials as a whole with the second five. The expected interaction emerged significant at the .01 level of confidence. The trend in means is displayed in Table 3. Highly motivated *S*s increased while *S*s of lower motivation decreased in attempted leadership with continued group experience.

Effect of difficulty. It was expected that participation would be greatest where problems were not so easy that little change was required, nor so difficult that everyone withdrew from the challenge. The results failed to support the hypothesis. The group data were examined by the five levels of problem difficulty ignoring practice. An F of 1.72 emerged when the interaction of difficulty and control variance was tested against the error within groups. (An F of 1.90 is significant at the .05 level for 8 and 168 *df*.) Groups with control variance of 0.00 or 1.44 showed less total participation on the easier problems, while groups with control variance of 0.64 did not. Table 4 summarizes the three related analyses of variance.

TABLE 1

PROPORTIONAL NUMBERS OF THE 51 GROUPS AND 255 MEMBERS TESTED UNDER VARYING CONDITIONS

A. Groups

Motivation Level	Control Variance				
	0.00	0.64	1.44	All	Ratio
High	9	9	9	27	9
Medium	4	4	4	12	4
Low	4	4	4	12	4
All	17	17	17	51	

B. Members

Motivation Level	Amount of Control							
	Control Variance 0.00	Control Variance 0.64			Control Variance 1.44		All	Ratio
	1X	1X	2X	3X	1X	4X		
High	45	27	9	9	36	9	135	9
Medium	20	12	4	4	16	4	60	4
Low	20	12	4	4	16	4	60	4
All	85	51	17	17	68	17	255	
Ratio	5	3	1	1	4	1		

TABLE 2

MEAN ATTEMPTED LEADERSHIP IN SECONDS TALKED PER PROBLEM BY THE AVERAGE MEMBER AS A FUNCTION OF MOTIVATION AND CONTROL VARIANCE

Motivation Level	Control Variance			
	0.00	0.64	1.44	All
High	25.2	28.4	34.7	29.5
Medium	28.6	18.9	18.2	21.9
Low	15.1	13.8	22.2	17.0
All	23.6	22.7	28.1	24.8

Analyses of Variance of Individuals Within Groups

The deviation of each member from his own group's mean was calculated for each problem. Since each group mean deviation was set at zero, all group differences, and the *simple* effects associated with motivation, control variance, practice, or difficulty were eliminated (although their interactions with amount of control were not). The variance in these deviations could be attributable only to amount of control of a member within a group, the interaction of amount of control with motivation, or period, or problem difficulty and the fluctuations between *S*s or within *S*s on successive trials.

TABLE 3

MEAN ATTEMPTED LEADERSHIP IN SECONDS TALKED PER PROBLEM BY THE AVERAGE MEMBER AS A FUNCTION OF GROUP PRACTICE AND MOTIVATION

Motivation Level	Amount of Practice		
	First Phase (Average of first five trials)	Second Phase (Average of second five trials)	All
High	28.2	30.8	29.5
Medium	22.1	21.7	21.9
Low	17.7	16.3	17.0
All	24.3	25.2	24.8

Two separate sets of analyses of variance were run. The first concerned the 17 groups with a control variance of 0.64 and the second concerned the 17 groups where variance in control was 1.44. Table 5 summarizes these two analyses. Amount of control did not vary in the remaining 17 groups where all members drew 1X, hence, these groups could have no remaining variance other than subject error. Nothing further was done with these groups.

As noted in Table 5, the variance estimate due to amount of control was significantly higher than the error between *S*s at the .01 level of confidence in both analyses. The ex-

TABLE 4

ANALYSES OF THE VARIANCE OF THE 51 GROUPS

Source	*df*	Variance	*F*	Source	*df*	Variance	*F*	Source	*df*	Variance	*F*
Motivation (M)	2	176,192.25	6.67**								
Control Variance (C)	2	31,745.75	1.20								
M × C	4	37,176.38	1.46								
Error between groups	42	25,403.17									
Between groups	50	32,630.29									
Periods (P)	9	2,327.11		Difficulty (D)	4	3,435.50		Phase (Ph)	1	2,823.52	
P × M	18	2,662.53		D × M	8	2,451.05		Ph × M	2	146,084.93	5.48**
P × C	18	3,509.14		D × C	8	4,471.80	1.72	Ph × C	2	42,868.13	
P × C × M	36	1,253.10		D × C × M	16	1,667.50		Ph × C × M	4	38,739.55	
Error within groups	378	3,242.01		Error within groups	168	2,599.40		Error within groups	42	26,739.53	
Within groups	459	3,055.82		Within groups	204	2,610.30		Within groups	51	32,524.47	
Total	509			Total	254			Total	101		

** $p < .01$.

TABLE 5

ANALYSES OF THE VARIANCE OF THE DEVIATIONS OF SUBJECTS FROM THEIR OWN GROUP MEANS

Source	17 Groups of Control Variance of 0.64			17 Groups of Control Variance of 1.44		
	df	Variance	*F*	*df*	Variance	*F*
Amount of control (A)	2	183,176.85	5.35**	1	497,431.00	11.03**
A × Motivation (M)	4	28,441.78		2	242,491.60	5.38**
Error between *S*s	76	34,534.32		79	45,090.02	
Between *S*s	82	37,862.55		82	55,421.05	
Period (P) × A	18	4,216.36	1.78*	9	5,196.97	1.25
P × A × M	36	2,787.73		18	6,161.12	
Error within *S*s	684	2,345.26		711	4,113.17	
Total	820	5,957.48		820	9,300.80	
Difficulty (D) × A	8	2,695.56	1.21	4	6,408.10	1.70
D × A × M	16	1,538.93		8	7,775.77	2.07*
Error within *S*s	304	2,268.34		316	3,760.60	
Total	410	9,367.05		410	14,196.86	

* $p < .05$.

** $p < .01$.

pected interaction of amount of control and motivation emerged in both situations but only in the more extreme 1X versus 4X condition did it reach significance at the .01 level of confidence. Table 6 shows the mean deviation of *S*s of varying amounts of control under different motivating conditions. Members of high control attempted more leadership than the others in their own group, particularly if all were highly motivated.

One unexpected result shown in Table 6 was the emergence of a "Vice-Presidential" control effect. Generally, attempted leadership was higher, the higher a member's amount of control, in an absolute sense. The average 1X members in the groups under the two control variance conditions deviated 0.84 and −2.42 seconds respectively from their own group means. The average deviation was 5.10 seconds for the 3X members and 9.68 seconds for the 4X members. However, 2X members, in competition with 1X and 3X members in groups with a control variance of 0.64, consistently attempted the least amount of leadership for all three motivation levels exhibiting an average deviation from their own group mean of −7.62 seconds.

Practice effect. Fig. 1 charts the trends in deviations of *S*s from their own group means with practice. An *F* of 1.78 significant at the .05 level was obtained for the interaction of period and status of 1X, 2X or 3X. A corresponding *F* of 1.25 (not significant) was found for the interaction of practice with 1X and 4X status. Inspection of Fig. 1 suggests that following an initially high rate of attempted leadership by persons of high control there was a drop followed by a rise ending with the 3X members attempting relatively as much leadership as the 4X members. The 4X members did not exhibit such change. It may be the 3X-2X-1X condition was more productive of conflict

TABLE 6

AVERAGE DEVIATION OF SUBJECTS FROM OWN GROUP MEAN IN SECONDS TALKED PER PROBLEM AS A FUNCTION OF AMOUNT OF CONTROL AND MOTIVATION

	Amount of Control				
	Control Variance 0.64			Control Variance 1.44	
Motivation	1X	2X	3X	1X	4X
High	1.61	−10.70	5.88	−4.66	18.66
Medium	−0.72	−6.62	8.78	0.36	−1.42
Low	0.69	−1.70	−0.38	−0.14	0.56
All	0.84	−7.62	5.10	−2.42	9.68

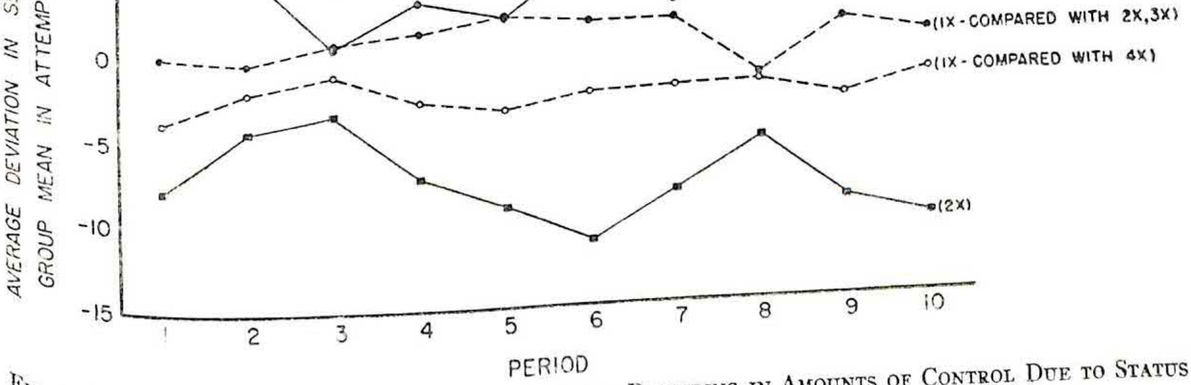


FIG. 1. TRENDS IN ATTEMPTED LEADERSHIP BY SUBJECTS DIFFERING IN AMOUNTS OF CONTROL DUE TO STATUS

and a struggle for status than the 1X-4X situation, but after some experience, the 3X member emerged attempting as much leadership as his 4X counterpart. Although motivation did not interact significantly with the above effects, inspection of the trends under different levels of motivation (not shown) suggested that under high motivation the decline and subsequent rise again of the 3X member was even greater than for all motivation levels combined.

Effects of Difficulty

Maximum attempted leadership by members of high control was expected with problems of average difficulty, problems not so easy that leadership was not required but not so difficult that everyone became apathetic.

As problems became more difficult, there was a tendency for the high control individuals to increase their tendency to attempt leadership relative to that attempted by the members of low control, although, as reported in Table 5, the trends were not significant. However, a significant interaction emerged in the 1X-4X situation. The interaction between difficulty level, amount of control and motivation was significant at the .05 level of confidence. The 4X members attempted more leadership at all difficulty levels under high motivation with an increase in this tendency as problems became more difficult. Under low motivation, this tendency was reversed with the 4X members tending to attempt more leadership than 1X members in easier problems and 1X members tending to display more attempted leadership on more difficult problems. The 3X members increased in attempted leadership as problem difficulty increased, regardless of motivation level.

SUMMARY AND CONCLUSIONS

Fifty-one groups of five *S*s each were studied on ten brief trials to examine the effects of differing motivation and amount of control in the tendency to attempt leadership in a group situation. On each trial, members rank-ordered judgments before and after group discussion. Groups varied in motivation to perform well according to pretest questionnaires and were assembled so that all members were either high, middle, or low in motivation to be accepted for Advanced ROTC training. Control, the ability to reward or punish other members, also was varied within each motivational group level. Attempted leadership was defined as the total amount of time spent by the *S* in talking in initially leaderless group discussions.

The results point to the significance of motivation and amount of control in the study of behavior in groups. Specifically, they suggest that a member's attempted leadership is higher, the higher his control and the higher the group's motivation. Control yields power only where members are highly motivated. Problem difficulty and group practice must also be considered in understanding the relations between power and leadership. On the whole, less leadership was attempted where problems were easy. With continued practice, highly motivated *S*s increased in attempted leadership while unmotivated *S*s did the opposite. Practice also had complex effects on attempted leadership as a function of control due to status.

REFERENCES

1. BASS, B. M. An analysis of the leaderless group discussion. *J. appl. Psychol.*, 1949, **33**, 527–533.
2. BASS, B. M. Outline of a theory of leadership and group behavior. *Tech. Rep.* No. 1, Contract N7 ONR 35609. Baton Rouge: Louisiana State Univer., 1955.
3. BASS, B. M., GAIER, E. L., FARESE, F. J., & FLINT, A. W. An objective method for studying group behavior. *Psychol. Rep. Monogr.*, 1957, **3**, 265–280.
4. LINDQUIST, E. F. Design and analysis of experiments in psychology and education. New York: Houghton Mifflin, 1953.
5. NOBLE, C. E. The meaning familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.
6. WILLERMAN, B. Organizational involvement as reflected in type of member complaint: An indirect method of measurement. *Tech. Rep.* No. 4, Contract N8 ONR 66216. Minneapolis: University of Minnesota, 1954.

Received May 31, 1957.

EFFECTS OF MOTIVATING INSTRUCTIONS ON REACTION TIME IN CEREBRAL DISEASE[1]

HAROLD L. BLACKBURN

State University of Iowa and VA Hospital, Iowa City

The influence of manipulable motivational factors on behavior in brain-damaged patients is a topic that has received little systematic investigation. Yet it is a problem with important implications. Among clinical workers, one finds two contrasting points of view (which are not always explicitly stated) regarding the significance of motivational variables in determining the nature and extent of behavioral impairment following cerebral injury or disease. The more traditional point of view conceives of the physical injury as being virtually the sole determinant of the behavioral deficits that are observed to follow the occurrence of the injury, and motivational or other factors are considered as being of little significance as determining variables (10, 15). Moreover, a contrast between this presumably anatomically grounded type of behavioral deficit and apparently similar behavioral deficits encountered in "functional" disorders (e.g., schizophrenia) is often made, the latter being conceived as being significantly if not primarily determined by motivational disturbances.

In contrast to the traditional idea that the physical injury is practically the sole determinant of the associated behavioral deficit in patients with cerebral disease, a more recent point of view conceives of the observed deficits following cerebral damage as being multiply determined and emphasizes the importance of special motivational procedures in rehabilitation work with these patients (4, 13, 18, 19). Standard procedures (orientation to the task with little or no urging or evaluation) are generally considered to be ineffective with such patients. On the other hand, procedures which involve failure experiences and the setting of difficult goals are frequently reported to result in "catastrophic reactions" with consequent severe disruption of performance. "Relaxing" procedures (emphasizing success experiences, reassurance, and minimal frustration) are generally considered to be optimally effective. However, since opinion to date regarding the effects of special motivational procedures has been based on nonquantitative clinical observations of complex and rather ill-defined performances, it seemed desirable to undertake systematic studies of the effects of motivational variables on the performances of brain-damaged patients in such a relatively simple and clearly defined task as the reaction time experiment.

A survey of the literature (1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 14, 16, 17) indicates that considerable doubt exists concerning the type of changes to be expected in the RT performances of either normal or brain-damaged *S*s, either as a function of simple practice or of special motivating conditions. The present study, which investigated the effects of three kinds of motivating instructions (standard, urging, relaxing) on choice reaction time in nonpsychotic patients with cerebral disease or injury and selected control patients, was designed to answer the following questions: 1. Do brain-damaged *S*s show changes in speed of choice reaction as a function of either of two types of special motivating instructions? 2. Are these motivational effects comparable in size and direction for both brain-damaged and control *S*s? Do brain-damaged *S*s show performance decrements under the standard instruction condition that are not shown by control *S*s?

METHOD

Subjects and Apparatus

A total of 195 patients from the University Hospital and the Veteran's Administration Hospital, Iowa City, were given two consecutive administrations of a test involving a choice reaction to one of two lights. The first administration was given always under standard instructions. The second administration was given under one of three sets of motivating instructions (standard, urging, relaxing). The patients were assigned on the basis of medical diagnosis to one of two groups: an experimental, or brain-damaged group, and a control group. One third of the patients in each diagnostic group were assigned to each instruction subgroup. In order to control for possible differential effects on the criterion scores of the factors of sex of patient or the facility in which he was hospitalized, the *S*s within each diagnostic group were further divided into subsamples. Assignment of *S*s was so made that the proportions of VA hospital males to University hospital males to University hospital females remained constant for each instructions-diagnosis combination.

---

[1] Condensed from a doctoral dissertation submitted to the State University of Iowa and done under the direction of Arthur L. Benton. This investigation was supported by a research grant (B-616) from the National Institute of Neurological Diseases and Blindness, of the National Institutes of Health, Public Health Service. Sincere thanks are due to Leonard Feldt for valuable statistical guidance and to Russell Meyers, Adolph Sahs, and Maurice Van Allen for their cooperation in obtaining and evaluating the patients used in this study.

The experimental group consisted of 78 nonpsychotic, brain-damaged patients who carried final and unequivocal diagnoses of supratentorial cerebral disease or injury. A variety of diagnoses was represented in the group, vascular disease and neoplasm being the most frequent. The control group consisted of 117 patients from the neurological, neurosurgical, and medical services who showed no evidence or history of cerebral disease or injury. A variety of physical complaints and diagnoses was represented in this group. The mean age of the 78 brain-damaged *S*s was 38 years; their mean educational level was 10 years. The mean age of the 117 control *S*s was 37 years, and their mean educational level was also 10 years. The general criteria for the selection of *S*s and the apparatus employed were identical with those employed in previous studies (2, 3).

Administration

The *S* was seated in front of the light panel of the apparatus. No arm rests were provided. The experimenter sat across the table on the control side of the apparatus. The following instructions were given:

> This is a test to see how fast you can react when you have to make a choice. This time, you place one hand on this key (point) and the other hand on this key (point) and rest them there without pressing down. I'll start a buzzer from over here again; and when the buzzer stops, *one* of the two lights on the panel in front of you will flash on. Keep watching for the light to flash, and when you are sure the right light is on, press down the right key, but *not* the left, as quickly as you can. When you are sure the left light is on, press down the left key, but not the right, again as quickly as you can. Never press down both keys at one time. It's either one or the other. Is that clear? (Repeat relevant portions as necessary). Now we'll try some practice (5 practice trials). That's the end of the practice; now we'll start testing (30 test trials).

Any response completed in .100 second or less, or any depression of the key on the incorrect side, regardless of whether the correct key was depressed before or after, was considered to be an anticipatory error, and an additional trial was given. Only times of correct responses were recorded. Thirty trials involving simple RT always preceded the choice RT procedure. The first 30 choice trials included 15 presentations of each light in a fixed, initially randomly determined order. The trials were given consecutively with no appreciable intervening interval. Following a rest period of approximately two minutes, one of the three following sets of instructions was given to precede a retest administration consisting of three warm-up and thirty experimental trials (15 presentations of each light in the reverse order of the initial administration presentations):

(*a*) *Standard or task-orienting instructions:* "Now let's try this choice test again. The same instructions apply as before. Let's have a few warm-ups first."

(*b*) *Urging and evaluating instructions:* "Now let's try this choice test again. The same instructions apply as before, only this time, I want to see just how *very* fast you can make the reactions. I've just figured your scores for the last series and your average time was ——. You should be able to do better than that. This time after each reaction, I will tell you in hundredths of a second how long it took you to react. Try to adjust your reactions so as to continually do better. Knowing how you are doing each time should help you. Let's have a few warm-ups first." (After each trial, *E* accurately reports the time and makes evaluative comments. Times equal to the average ± .03 second elicit 'OK'; slower times elicit 'poor' or 'too slow'; faster times elicit 'good' or 'fine.' In each case, *S* is urged to do even better.)

(*c*) *Relaxing and reassuring instructions:* "Now let's try this choice test again. The same instructions apply as before, only this time, I think you would do better if you just *relax*. You've been doing fine so far but just *relaxing* seems to help everybody. Still try to react as quickly as possible but don't concern yourself about how you are doing. Relax as much as possible while you are waiting for the light. Let's have a few warm-ups first." (After practice, assure *S* he is doing fine—even better than before—and should just continue to relax. Repeat assurance after trials 5 and 20 or whenever *S* makes spontaneous inquiry.)

RESULTS

Initial scores. A preliminary analysis of initial test scores (sum of the time measures for the first 30 trials under standard conditions) was performed in order to evaluate any differences in absolute level of performance before instructions treatments were imposed. Significant initial differences ($F = 31.62$, *df* 1 and 189, $p < .001$) in absolute level were found between brain-damaged and control *S*s, the brain-damaged *S*s being consistently (and, in most instances, markedly) slower than the corresponding control *S*s.

In view of the differences associated with diagnostic group, the effects of sex and hospital residence on initial level were evaluated by performing two separate 3 x 3 factorial analyses of variance, one for the brain-damaged and one for the control *S*s. Within each diagnostic group, a significant difference among the subpopulations ($F = 4.02$, *df* 2 and 69, for brain-damaged, $F = 4.72$, *df* 2 and 108, for control; $p < .05$ in each case) was found. In both groups, the Veterans Hospital males were the fastest, the University Hospital males were next fastest, and the University Hospital females were the slowest.

The same analyses also provided an opportunity to examine the relative size of the differences among the instructional groups which resulted from the random assignment of *S*s to the several instructions treatment groups. No noteworthy difference in mean initial score was found among the three instructional subgroups of control *S*s. However, contrary to expectations based on the *random* assignment of the *S*s to instructions subgroups, an initial difference ($F = 3.63$, *df* 2 and 69, $p < .05$) was found among these three subgroups of the brain-damaged population. No satisfactory explanation of this finding can be offered except in terms of an unusual chance happening.

To meet the difficulties associated with multiple differences in initial level of performances, an analysis of covariance of the retest scores (sum of the time measures for the second 30 trials for each *S*), with initial scores being used as the control variable, was considered. However, the assumptions for such an analysis could not be met due to a marked heterogeneity of regression within cells.

Gain scores. Since the assumptions for an analysis of covariance could not be met, gain scores (the sum of the 30 time measures *before* instructions treatments minus the sum of the 30 time measures *after* instructions treatments) were utilized as the criterion measure.

Criterion score means and their standard deviations for each of the 18 cells are presented in Table 1. Inspection of this table reveals marked heterogeneity of variance from cell to cell. Application of Bartlett's test for homogeneity of variance statistically confirms the judgment that these variances differ by more than chance ($p < .001$). However, empirical findings, as discussed by Lindquist (12), indicate that the F test of analysis of variance is not very sensitive to any failure to meet the assumptions of normality of distribution or homogeneity of variance. Moreover, means and variances of the 18 cells were not correlated ($r = .01$). Therefore, it was decided that a 3 x 2 x 3 factorial analysis of variance of these criterion measures would permit at least tentative conclusions as to the significance of the various main effects and interactions. In this respect, it should be noted that because of the heterogeneity of variance, the true risk of a Type I error (rejecting a true null hypothesis) is, by the best available estimates, about two to four per cent more than the apparent level of significance.

Three specific null hypotheses were posed with respect to the criterion score means of the treatment populations corresponding to the several diagnostic-instructional categories: (*a*) the effect of instructional differences upon the criterion score means is identical for the two diagnostic groups; (*b*) in the combined diagnostic populations, there is no difference in criterion score means as a function of differences in motivating instructions; (*c*) in the combined instructions populations, there is no difference in criterion score means as a function of differences between the diagnostic populations.

An examination of the subpopulations factor (Veterans Hospital males, University Hospital males, University Hospital females) and its interactions does not bear directly on any of the three hypotheses and is entirely secondary to the purposes of the present investigation. However, restriction of sampling to any one of these subpopulations would have seriously limited both the available pool of brain-damaged *S*s and the generality of the present

TABLE 1

MEAN CRITERION SCORES (IN .01 SEC.)
AND THEIR STANDARD DEVIATIONS
BY SUBGROUPS

Population		Instructions Subgroup		
		Standard	Urging	Relaxing
Brain-damaged Subpopulation				
VA males	M	−49	116	25
	SD	69	159	97
Univer. males	M	−22	112	17
	SD	50	53	255
Univer. females	M	−37	172	50
	SD	81	75	66
Control Subpopulation				
VA males	M	24	86	−22
	SD	61	41	76
Univer. males	M	−26	75	10
	SD	76	120	87
Univer. females	M	31	86	69
	SD	63	67	88

findings. Since the statistical tests rendered tenable the assumptions of no subpopulations effect ($F = 1.95$, *df* 2 and 177, $.20 > p > .10$) on the criterion scores and no double or triple interaction in the populations involving this factor ($F < 1.00$; $p > .50$ in each case), subsequent tests of the experimental factors (instructions and diagnosis) and conclusions with respect to these factors were made possible at more general levels, unrestricted to the individual subpopulations.

The first null hypothesis is, in effect, an assertion of zero interaction between diagnostic category and instructions. There is some question whether the significance of the F test ($F = 3.53$, *df* 2 and 177, $p \approx .03$) for this interaction should be attributed to an intrinsic interaction in the population or to failure of the data to meet the assumption of homogeneity of variance or both. However, since the empirical studies of heterogeneity of variance indicate that an F of this size is extremely close to a true five per cent level for the interaction hypothesis, it would seem advisable that the effects of instructions and diagnosis be studied separately at each level of the other and that the observed differences be evaluated with respect to error terms appropriate for and computed from each limited segment of the data.

Mean gain scores for each of the six diagnosis-instructions treatment combinations are presented in Table 2. From inspection, it will be noted that marked differences in mean criterion scores among the three instructions treatment groups were found, both within the brain-damaged population and within the control population. Repetition of standard instructions appeared to result in a decrement in performance for brain-damaged *S*s and a minimal increment for control *S*s. The introduction of urging instructions appeared to result in marked improvements over initial test scores for both groups. The introduction of relaxing instructions appeared to have no differing effects from repetition of the standard instructions for control *S*s. The brain-damaged *S*s, however, showed an apparent improvement in performance under the relaxing condition as compared with their previously noted decrement under the standard condition.

TABLE 2

MEAN GAIN SCORES FOR EACH OF THE SIX DIAGNOSIS-INSTRUCTIONS TREATMENT COMBINATIONS

Diagnoses	Instructions Group Standard	 Urging	 Relaxing
Brain-damaged	−38	127	28
Control	10	82	9

Separate analyses within each diagnostic group confirmed the significance of the observed main instructional effects ($F = 11.65$, *df* 2 and 75, $p < .001$, for brain-damaged group, $F = 10.50$, *df* 2 and 114, $p < .001$ for control group). Analysis of the simple instructional effects showed that the urging instruction differed significantly in its effects from each of the other two instructions (standard and relaxing) within both the brain-damaged and the control populations. The standard instructions and the relaxing instructions did not differ significantly in their effects within either population. With respect to the control *S*s, the observed difference in criterion scores between the standard and relaxing instructions group was minimal and nonsignificant. On the other hand, the observed difference with respect to brain-damaged *S*s was relatively large and approached significance at the pre-established level ($.10 > p > .05$).

Under each instructions condition, the brain-damaged *S*s showed apparently greater changes than the control *S*s. Under the standard condition, these changes were in opposing directions (*vide supra*), the difference in mean criterion scores for the two diagnostic groups being statistically significant ($F = 7.28$, *df* 1 and 63, $p < .01$). Under the other two conditions, the amount of improvement among the brain-damaged *S*s was not demonstrably different from that shown by the control *S*s, although the observed difference with respect to the urging condition approached significance ($F = 3.18$, *df* 1 and 63, $.10 > p > .05$ for the urging condition; $F = 0.39$, *df* 1 and 63, $p > .50$ for the relaxing condition).

In view of the marked differences between the brain-damaged and control *S*s with respect to initial level of performance, the question arises as to whether the observed differences in gain scores between the two diagnostic groups are a function of these differences in initial level. The question is of further significance in view of the observed differences in initial level among subsamples in both groups and among the instructional groups within the brain-damaged population.

In order to analyze the relation between initial and gain scores, the *S*s in each of the six treatment combinations were divided into "fast" and "slow" initial performers on the basis of their relationship to the median for that particular cell. The subsamples were pooled for purposes of this analysis. Separate analyses of variance of each of these six differences in criterion scores indicated that none of the observed differences attained statistical significance. Therefore, it may be inferred that if any relationship exists between initial and gain scores, the effect of initial level or subsequent gain scores is of little relative importance when compared with the effects of other factors influencing level of subject performance.

Blocks of trials. The finding of significant effects of instructions for both the brain-damaged and control groups was based on an analysis of gain scores that were derived from total performance over a considerable number of trials. The question arises as to whether these instructional effects are operative throughout the course of performance over the 30 postinstructional trials or only during part of it, e.g., the first few trials. To answer this question, the mean time scores per trial for blocks of 10 trials each for both the preinstructional and instructional conditions and for each diagnosis-instructions treatment combination were computed.

With respect to the preinstructional condition, it was found that the control *S*s showed no evidence of either positive or negative practice effects and that the brain-damaged *S*s showed a slight negative practice effect. With respect to the several postinstructional conditions, the course of performance was very close to a horizontal line in every case, indicating that whatever the instructional effect might have been, it operated fairly constantly throughout the course of the 30 trials.

DISCUSSION

The present study was concerned primarily with the RT responsiveness of nonpsychotic patients with supratentorial disease or injury of the brain to three types of motivating instructions. The performances of those *S*s were compared with those of nonpsychotic medical patients with no evidence or history of cerebral disease or injury, but who were suffering from one type of illness or another. It was believed that, for purposes of an initial study in the area, such *S*s constituted a more appropriate control group than healthy normals in that the possible influence of the factors of illness and hospitalization was controlled.

Initial scores. Significant differences were found in initial level of performance between brain-damaged and control *S*s. As was expected, brain-damaged *S*s were clearly and markedly slower as a group than control *S*s. However, analyses failed to reveal any significant relation between initial and gain scores.

Gain scores. Analysis of gain scores with respect to their relationships with diagnosis and instructions constituted the primary analysis of this study. On the basis of the finding that a significant diagnosis-instructions interaction existed, it was decided that analysis of the simple effects of these two primary experimental factors was most appropriate. With respect to these simple effects, it was found that the two diagnostic groups responded differently to the standard instructions conditions, with brain-damaged *S*s showing a performance decrement postinstructionally and control *S*s a performance increment. This observed difference between diagnostic groups was significant. The direction of these trends in performance with practice is consistent with previous studies and clinical impressions. Moreover, these findings support the suggestion derived from the findings of an earlier study (2) that a larger number of trials than previously employed (60 vs. 30) might reveal a difference in practice trends which would be significant.

Both diagnostic groups showed apparent performance increments as a function of relaxing instructions. The increment shown by the brain-damaged *S*s was apparently greater than that shown by control *S*s, but not significantly so. Among control *S*s, the observed difference in mean criterion scores between the standard and relaxing instructional subgroups was minimal (10 vs. 9) and nonsignificant. Among brain-damaged *S*s, the same difference was sizeable (−38 vs. +28) and approached significance, and the *direction* of this difference (a decrement under the standard condition and an increment under the relaxing condition) is consistent with clinical contentions that special motivating conditions which assure success and which foster a relaxed attitude in the patient will improve the performance of brain-damaged *S*s and will tend to counteract the usual practice decrements seen in these *S*s under standard conditions. However, since the observed difference was not demonstrably significant, it provides only minimal support for such an idea. In any case, the present findings suggest that such procedures are at least equally as effective as standard procedures.

Urging instructions elicited the greatest improvements in performance, both for brain-damaged and for control *S*s. The increment shown by brain-damaged *S*s was again apparently greater than that shown by control *S*s but not significantly so. Within each diagnostic group, the effects of urging instructions were significantly different from the effects of either of the other conditions. With respect to control *S*s, this finding is consistent with some earlier investigative conclusions (11, 16). With respect to brain-damaged *S*s, the present finding conflicts with earlier reports that attempts to improve the performance of such *S*s through motivational manipulations were unsuccessful (10, 15). Moreover, it is not consistent with the clinical impression that brain-damaged *S*s frequently react catastrophically to failure instructions and the setting of difficult goals. Only 3 (12 per cent) of the 26 brain-damaged *S*s in this instructional category showed performance decrements and, in each case, the decrements were relatively small. Moreover, 4 (10 per cent) of the 39 control *S*s in the same instructional category showed similar decrements, suggesting that brain damage is not the crucial subject variable which determines such decrements. The present data force the conclusions that brain-damaged *S*s benefit from complex instructions involving both failure and the setting of difficult goals and that they benefit from these instructions to at least the same degree and in the same proportions as control *S*s.

Clinical implications and limitations. In general, the instructions conditions were so selected as to reflect or approximate the main motivational procedures employed in the clinic. The standard condition closely parallels the typical test situation in that the *S*s are instructed to do their best at the beginning and no knowledge of results, evaluation, or further encouragement is given during the course of testing. It may be noted that this condition, which is typical of diagnostic test procedures, provided the best diagnostic discriminations and that discrimination was enhanced by including additional trials (i.e., 60 vs. 30).

The relaxing instructions were so selected as to reflect what might be called the current consensus of opinion among clinicians responsible for the retraining of skills in brain-damaged *S*s. *S*s were told that they had succeeded initially. Then the suggestion was made that they would probably do even better. After further attempts, they were reassured that they had met the new goals set for them. Minimal frustration, success experience, and reassurance were the intended major elements of this procedure. Failure to achieve clearly significant effects from such a procedure suggests at least two hypotheses. First, it may be that interfering intensities of concern were not aroused by the type of task employed in this investigation so that procedures designed to reduce concern or anxiety would not be expected to improve performance. Lack of concern in the present task may be related either to insufficient opportunity for the naive *S* to evaluate his own RT performance or to the relatively unimportant nature of the task. This would be contrasted, then, with the usual clinical situation in which the aphasic patient, for example, is fully and painfully aware of his failures and highly conscious of the importance of the skill in which he is deficient. Under such conditions, relaxing procedures might prove to be efficacious.

A second hypothesis which would account for the present findings with respect to relaxing instructions is suggested both by the data and by impressions of the experimenter. The relaxing instructions were associated with the greatest interindividual variability within both diagnostic groups, suggesting that the particular instructions employed may have been ambiguous in character. This suggestion is supported by subjective observations to the effect that many *S*s construed these directions to mean "lie back in your chair and press the key casually." Other *S*s appeared to understand and to respond to the intended meaning of the instructions.

Whatever the essential differences in the experimental and clinical conditions with respect to relaxing instructions may be, the present data fail to lend any substantial support to current clinical preference for this type of motivational procedure. Particularly is this true when the results with respect to the urging instructions conditions are considered. Admittedly, the urging condition in the present investigation was complex. It included an initial failure experience in that *S*s were told that they should have done better on the first 30 trials. It included factual knowledge of results so that some postinstructional trials led to failure experience and some to success experience. However, for most *S*s in both groups, the predominant experience was success since about 90 per cent of all *S*s improved in performance. Moreover, this condition included the setting of relatively difficult goals ("do better each time") and continual evaluation and urging by the examiner.

In many respects, then, the urging instructions condition involved factors or elements which have been observed or claimed to precipitate catastrophic reactions and severe decrements in performance among brain-damaged *S*s. The present findings that brain-damaged *S*s show their only demonstrable improvement under this condition and that they improve to at least the same degree and in the same proportions as control *S*s are in direct contrast to the clinical observations or impressions and raise some interesting and serious questions concerning the selection of motivational procedures in rehabilitative work with such *S*s. One possible hypothesis, of course, is that the present results are again a function of the nature of the task. It may be that attempts to raise the motivational level of *S*s who are already concerned or anxious about their handicaps (e.g., aphasics or hemiplegics) would lead to performance decrements or disorganized and competing behaviors. Another possible hypothesis is based on the fact that, in the present study, most *S*s achieved relative success in meeting the experimenter-defined goals. It may be that the setting of goals which are beyond the current capacities of the patient would produce interfering responses and consequent decrements in performance. In any case, the present data provide no support for a point of view that warns against too vigorous "pushing" of brain-damaged patients and little support for one that emphasizes the desirability of relaxation and reassurance procedures.

The present study has been concerned with performance on a simple, high speed task, which is minimally affected by simple practice. Discrepancies between the present experimental findings and earlier clinical impressions suggest the need to investigate possible interactions between motivational instructions and the nature of the task. More complex, important, ego-involving, and anxiety-producing tasks might be employed. Moreover, the complex instructions of the present investigation might be broken down into simpler elements which could be studied separately. The influence of knowledge of results, of success experiences, failure experiences, evaluation, reassurance, and urging might then be determined independently of their joint effects and interactions. Some further attempt might also be made to reduce the suspected ambiguity of the relaxing instructions which were employed. Finally, a determination of the effects of these instructional procedures among healthy *S*s might be attempted in order to assess the influence of such factors as illness and hospitalization, the possible effects of which were controlled in the present study.

SUMMARY

The main purpose of the present study was to determine the RT responsiveness of hospitalized, nonpsychotic patients with supratentorial cerebral disease or injury to three types of motivating instructions and to compare their performances with those of a similar group of hospitalized medical patients without cerebral disease. Initially, 30 visual choice RT trials were administered to each *S* under standard conditions. This was followed by differential instructions (standard, relaxing, urging) and by 30 postinstructional trials identical with the initial 30 trials. The criterion score for each *S* was the difference in total time scores between his 30 preinstructional and 30 postinstructional trials.

Marked initial differences in RT between brain-damaged and control *S*s were found.

However, initial differences were not related to criterion scores within any of the six diagnosis-instructions combinations. Practice effects prior to instructions were minimal. Changes over blocks of trials following instructions were also minimal.

Standard instructions resulted in a decrement for brain-damaged *S*s and an increment for control *S*s. Both the relaxing and urging instructions resulted in increments for both groups, with brain-damaged *S*s showing the apparently, but not significantly, greater improvement under both conditions. Within each diagnostic group, the urging instructions were clearly and significantly superior to the other two instructions in effecting improvements in performance. Relaxing and standard instructions were not demonstrably different in their effects within either group. However, there was a suggestion that relaxing instructions might be more effective than standard instructions with brain-damaged *S*s.

REFERENCES

1. BAXTER, B., & TRAVIS, R. C. The reaction time to vestibular stimuli. *J. exp. Psychol.*, 1938, **22**, 277–282.
2. BENTON, A. L., & BLACKBURN, H. L. Practice effects in reaction time tasks in brain-injured subjects. *J. abnorm. soc. Psychol.*, 1957, **54**, 109–113.
3. BLACKBURN, H. L., & BENTON, A. L. Simple and choice reaction time in cerebral disease. *Conf. Neurol.* 1955, **15**, 327–338.
4. DUNBAR, H. F. *Emotions and bodily changes.* New York: Columbia Univer. Press, 1954.
5. EVANS, J. E. The effect of distraction on reaction time. *Arch. Psychol.*, 1916, **5**, No. 37.
6. FARBER, I. E., & SPENCE, K. W. Effects of anxiety, stress, and task variables on reaction time. *J. Pers.*, 1956, **25**, 1–18.
7. GOLDSTEIN, K. *After effects of brain injuries in war.* New York: Grune & Stratton, 1942.
8. HENMON, V. A. C., & WELLS, F. L. Concerning individual differences in reaction times. *Psychol. Rev.*, 1914, **21**, 153–156.
9. HOVLAND, C. I. The influence of adaptation illumination upon visual reaction time. *J. gen. Psychol.*, 1936, **14**, 436–359.
10. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. II. New York: Ronald Press, 1944. Pp. 971–1032.
11. JOHANSON, A. M. The influence of incentive and punishment upon reaction time. *Arch. Psychol.*, 1922, **8**, No. 54.
12. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
13. LONGERICH, M. C., & BORDEAUX, J. *Aphasia therapeutics.* New York: Macmillan, 1954.
14. MILES, W. R. The reaction time of the eye. *Psychol., Monogr.*, 1936, **47**, 268–293.
15. NICHOLS, I. C., & HUNT, J. McV. A case of partial bilateral frontal lobectomy. *Amer. J. Psychiat.*, 1940, **96**, 1063–1083.
16. ROSENBAUM, G. Stimulus generalization as a function of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
17. SMITH, K. U. Bilateral integrative action of the cerebral cortex in man in verbal association and sensorimotor coordination. *J. exp. Psychol.*, 1947, **37**, 367–376.
18. STRAUSS, A. A., & KEPHART, N. C. *Psychopathology and education of the brain-injured child.* Vol. 2. New York: Grune & Stratton, 1955.
19. WEPMAN, J. M. *Recovery from aphasia.* New York: Ronald Press, 1951.

Received June 3, 1957.

AUTONOMIC FEEDBACK: THE PERCEPTION OF AUTONOMIC ACTIVITY[1]

GEORGE MANDLER, JEAN M. MANDLER, AND ELLEN T. UVILLER
Harvard University

During the past several years there has been an increasing concern with the measurement and manipulation of the anxiety variable in human subjects. This interest is in contrast to May's summary in 1950: ". . . the problem of human anxiety has been absent from the strictly experimental specialties of psychology . . . (and) . . . has largely been omitted from the other branches of academic and theoretical psychology as well." (8, p. 99).

The measurement of anxiety has largely relied on the use of self-rating scales such as the Manifest Anxiety Scale (10) and the Test Anxiety Questionnaire (7). These scales have depended for their construction on some a priori definition of anxiety. In the case of the Manifest Anxiety Scale, this definition leaned heavily on the identification of symptoms consonant with Cameron's specification (2). For the Test Anxiety Questionnaire, the defining operations for item choices were even more vaguely dependent on an intuitive appreciation of the anxiety concept. The point need not be labored that these item choices depend to a large extent on the rather nebulous agreement as to what anxiety "really is." Such an agreement would be easily reached if the anxiety concept were securely anchored in a theoretical framework which provided invariant indices. Such indices are not available for anxiety as an individual difference variable in human subjects. Neither psychoanalytic theory nor the neo-Hullian formulation, to cite the two most frequently used models, provide clear-cut definition for the identification of the variable in human subjects.

This lack of specification for a variable which is subsumed under the emotions is not a novel problem in psychology. In 1890, William James issued a plea for a unified theoretical basis for emotional variables:

> But as far as 'scientific psychology' of the emotions goes, I may have been surfeited by too much reading of the classic works on the subject, but I should as lief read verbal descriptions of the shapes of the rocks on a New Hampshire farm as toil through them again. They give one nowhere a central point of view, or a deductive or generative principle. They distinguish and refine and specify *in infinitum* without ever getting on to another logical level (3, p. 448).

More recently Brown and Farber (1) have suggested a lucid solution for this problem within a Hullian framework. Their formulation, however, is not intended to make the definition of any particular emotional variable more specific. It appears, for the present, to be valuable to specify objective variables which are independent of any general solution for a theory of the emotions. In particular, it is suggested that the concept of anxiety has frequently been used in a global and complex fashion, subsuming many different constituent variables.

The study to be presented here is the first in a series of investigations of one such variable, that of autonomic feedback. Autonomic feedback is defined as the relationship between autonomic response and the subject's reported perception of such response-induced stimulation. While this feedback phenomenon encompasses a variety of different visceral as well as skeletal effects, the major source of stimulation is associated with the activity of the autonomic nervous system. We suggest that this phenomenon represents one variable which has frequently been described as part of the anxiety complex. We will take a position of theoretical neutrality on its specific relationship to anxiety. The study of autonomic feedback may be theoretically useful and empirically productive whether or not it is related to the anxiety concept. However, the three major sources of this concept do stem from general considerations of emotions and anxiety:

1. Autonomic feedback is related to James' position that so-called emotional states can best be described as perceptions of bodily (physiological) activities. It is also related to Lange's (and more recently Wenger's) (4, 11) definition of emotions as patterns of physiological response.

---

[1] This study was supported by a grant from the Foundations' Fund for Research in Psychiatry. It was reported at the April, 1957 meetings of the Eastern Psychological Association.

2. In the psychoanalytic literature, insofar as it has stayed close to Freud's original formulations, anxiety is primarily a perceptual or ego phenomenon. It is the subject's report of a subjective state, rather than the objective occurrence of any symptoms, which defines anxiety.

3. The anxiety scales mentioned above contain an appreciable number of items related to the subjective report of physiological events.

The first question to be raised concerning the problem of autonomic feedback is a simple one. What is the relationship between the occurrence of autonomic response and the subject's reported perception of such activity?

Physiological events are, in the daily life of the human organism, private events. They stimulate receptors, primarily kinesthetic ones, and are therefore one of the prior conditions for the subject's report of their occurrence. However, the report of the perception of these events is necessarily couched in words which have a different history than do those words which report the perception of public, external events. Skinner (9) has presented several arguments for the nonacceptability of such emotional terms for psychological science; they are primarily based on an analysis of the conditions of reinforcement for such words as "anxious," as compared with the conditions of reinforcement for such words as "red" or "table." Briefly, the contingency between the public event or object and the occasions of reinforcement for the terms applied to them is under the control of the reinforcing community. Such control is not possible for the acquisition of mental or emotional terms. A child will be reinforced for the use of the term "table" whenever in the presence of a table; he may be reinforced for the use of the term "anxious" or "angry" under a variety of internal stimulating conditions which are unknown to the reinforcing community. The result is a wider variability in the use of the latter as compared with the former class of terms.

Under such conditions, it is unwise to expect a simple relationship between the observation of autonomic activity—once made public—and the subject's report of its occurrence. In this initial study we use a twofold approach to the problem of report, and then, given a presumably reliable measure of report, we relate it to autonomic activity.

An examination of the relationship between self-report and autonomic activity is also relevant to the validity of one of the assumptions underlying the anxiety scales mentioned above. It is often tacitly assumed that the report of these subjects is veridical. That is to say, a subject who reports a high degree of subjective anxiety or of symptoms indicative of such anxiety does in fact exhibit these symptoms. However, two different states of affairs may account for a report of considerable internal response: (*a*) The high scoring subject may actually exhibit a greater amount of symptoms or autonomic activity than the low scoring subject, or (*b*) the high scoring subject may not differ from the low scoring subject in reactivity but be more sensitive to and more likely to report small changes in autonomic response.

This study will attempt to deal with this issue by examining the level of autonomic reactivity in two groups of subjects, those who report a high as against a low degree of autonomic feedback.

METHOD

Design. A questionnaire dealing with the frequency and intensity of autonomic self-perceptions was administered to 166 *S*s. The 19 highest scoring *S*s and the 13 lowest scoring *S*s were exposed to an intellectual stress situation. Subsequently, they were interviewed as to their autonomic perceptions during the stress experience and ranked on the basis of this interview. The records of the 14 *S*s who obtained high scores on both the questionnaire and the interview and the 9 *S*s who scored low on both instruments were then examined for level of autonomic reactivity.

Questionnaire construction. The Autonomic Perception Questionnaire (APQ) consists of three sections. The first section requires free response descriptions by the *S*s of their states of feeling and reactions when (*a*) in a state of anxiety and apprehension and (*b*) in a state of pleasure. The *S*s were instructed that when later questions referred to anxiety or pleasure, they should understand by those terms the state which they described in this section. The second section consists of 30 graphic scale items, each of which deals with the perception of bodily activity. These items cover seven areas of bodily reaction: heart rate, perspiration, temperature changes, respiration, gastrointestinal disturbance, muscle tension, and blood pressure. An item is composed of such questions as: "When you feel anxious, how often are you aware of any change in your heart action?" with a 14.5 cm scale, the end points of which are marked "Never" and "Always." *S* can mark the scale at any point. For scoring purposes, the scale is divided into 10 equal intervals, and the *S* is assigned a score from zero to nine for each item. Of the 30 self-rating scales, 21 relate to feelings when in a state of anxiety, and nine relate to the state of pleasure.

The third section of the questionnaire consists of 70 MMPI items. Fifty are from the Manifest Anxiety Scale, and an additional 20 were selected from the MMPI as dealing with reports of internal bodily stimulation. These 20 items, together with 14 contained in the MAS, make up a new 34-item Body Perception Scale (BPS).

The questionnaire was administered to 166 Harvard College students (primarily sophomores) as part of a course requirement. Prior to the administration of the questionnaire, *S*s were asked to indicate whether they would be willing to volunteer for subsequent sessions, without being informed of the nature of these sessions.

The stress situation. The *S*s for the stress experiment were selected on the basis of their total score on the 21 anxiety items of the APQ. Of the 25 highest scoring and the 25 lowest scoring *S*s who were requested to participate in another individual session, 19 Highs and 13 Lows kept their appointments. Volunteering did not affect the selection of *S*s and showed no differential tendency in participation.

The stress situation consisted of three tasks with concomitant physiological recording. *S* was seated with his back to the experimenter and the recording apparatus, facing a wall on which the task materials were presented. After a 20-minute adaptation period, *S* was told that he was to be given a series of standardized intelligence tests which should present no difficulty to the average college student. The tasks, a logical reasoning test, a vocabulary test, and a verbal maze, were designed to be extremely difficult, and no *S* was able to complete any of them satisfactorily. They were presented to all *S*s in the same order:

1. The reasoning test consisted of 15 items. *S* was presented with a series of propositions. He was then given 15 statements and was required to indicate the truth or falsity of these deductions from the propositions. The mean number correct on this test was 7.2 (out of 15).
2. The vocabulary test consisted of 10 items, each of which presented a word and a choice of five definitions. The mean number correct was 2.6 (out of 10), which indicates the difficulty of the task.
3. In the Verbal Maze, *S* was presented with a series of choice points consisting of 2, 3, 4, or 5 letters. Each particular combination of letters had one correct choice which permitted *S* to go on to the next choice point. A trial consisted of 12 such choice points, which were repeated for every trial. After four trials, *S* was told that he was doing very badly (which every *S* was in fact) and then was given two more trials.

Physiological measures. A six-channel Grass polygraph was used to obtain measures of autonomic activity.[2] The basic equipment consisted of a Grass pen-recorder, driven by standard Grass D.C. amplifiers. Five channels were used in the present study:

1. *Heart rate.* A cardiotachometer gives either a direct reading of the conventional EKG record, or a tachometer record which supplies a reading of interbeat intervals which can be directly translated into beat-to-beat changes in heart rate.
2. *Psycho-galvanic response.* Preamplification was supplied by a Fels Dermohmmeter,[3] which incorporates an automatic range reset feature. The subject current is 70 microamps. Cup electrodes were used on the *S*'s palms.
3. *Respiration.* A standard pneumograph and a Grass PT-5 pressure transducer were used.
4. *Face temperature.* A Tele-Thermometer[3] with a thermistor probe placed on *S*'s cheek was employed.
5. *Blood volume.* A Waters oximeter, measuring changes in the opacity of the earlobe, was used as a pickup. The major variable contributing to change in opacity is assumed to be blood volume, which in turn is related to pressure changes.

Postexperimental interview. Following the stress situation, the *S* was given a structured interview to elicit his perceptions of autonomic changes during stress. Eight questions tapped areas similar to those used in the Autonomic Perception Questionnaire. The interview was scored by assigning values of two, one, and zero to the *S*s' responses, corresponding to reports of much, little, or no autonomic feedback.

RESULTS AND DISCUSSION

Questionnaire and interview data. Table 1 shows the intercorrelations between the two

---

[2] Major credit for the instrumentation goes to Albert Grass and Walter Bird of the Grass Instrument Company, Quincy, Massachusetts.

[3] Obtained from the Yellow Springs Instrument Co., Yellow Springs, Ohio.

TABLE 1

INTERCORRELATIONS AND MEANS FOR QUESTIONNAIRE SCALES

	Anxiety APQ	Pleasure APQ	MAS	BPS
Anxiety APQ	—			
Pleasure APQ	.50 .45	—		
MAS	.52 .51	.31 .28	—	
BPS	.52 .52	.20 .37	.67 .73	—
Mean volunteers ($N=74$)	81.6	25.7	17.51	6.46
Mean nonvolunteers ($N=92$)	81.6	22.8	16.34	6.24

Note.—The first correlation given is for volunteer *S*s, the second for nonvolunteer *S*s.

parts of the APQ (Anxiety and Pleasure), the MAS (Manifest Anxiety Scale) and the BPS (Body Perception Scale). All correlations, except that between Pleasure APQ and BPS for volunteers, are significant at the .01 level or better. The highest correlation is that of .70 between the MAS and the BPS. When the 14 overlapping items are removed from the BPS, the correlation drops to .60. The hypothesis that the Manifest Anxiety Scale is significantly related to *S*s' report of internal perceptions is sustained. The Anxiety APQ, the major concern in this investigation, shows consistent correlations around .50 with the three other scales. This finding supports the view that the questionnaire has some communality with other anxiety-related measures. Of special interest are the positive correlations shown by the Pleasure APQ, indicating that *S*s who report more internal stimulation do so whether the reference is to pleasurable or to anxiety-arousing situations. Whether this is an autonomic or a perceptual phenomenon or a function of response set is, of course, open to question. Finally, there are no significant differences between volunteers and nonvolunteers on either correlations or scale means. This may be related to the fact that *S*s volunteered before being aware of the nature of the tests to be taken.

As indicated above, the 19 highest scoring and the 14 lowest scoring *S*s on the Anxiety APQ participated in the stress situation. The range of their postexperimental interview scores extends from 0 to 16, which is the same as the theoretical range. The mean score for the high APQ *S*s was 7.2; for the low APQ *S*s, it was 4.1 ($p < .05$). This finding suggests some validity in the APQ, which taps the *S*s' general perception of autonomic reactions. The specific report of autonomic reactions in a stress situation is related to the more general APQ score.

The interview scores of the 32 *S*s were divided at the median. Fourteen of the 19 high APQ *S*s were also high on the interview rating, while nine of the 13 low APQ *S*s were also low on the interview rating. The 23 *S*s who showed concordance in their questionnaire and interview scores were selected for the following analysis of autonomic responses.[4]

Physiological findings. For purposes of analysis, the polygraph record was divided into ten reaction periods which represented various aspects of the test process: initial instructions, instruction periods for each test (three periods), and test presentations (six periods). The choice of measures of autonomic activity was dictated by two considerations. The first was previous utility in other studies. Since there are no such measures as the *best* indicator of changes in heart rate, PGR, etc., it was considered to be most efficient to describe the reactivity on any one channel of autonomic activity in terms of more than one measure. The second consideration was dictated by theoretical reasons. Since our primary interest was in autonomic perception, we attempted to select those response aspects which are most likely to be the basis of perceptual discriminations. Such indices as total increase in autonomic response and the variability of responses were considered to fulfill this criterion. The following measures were used:

Heart rate:

1. The mean autonomic lability score[5] for the first and last reaction periods.
2. The variability of mean heart rate over all ten periods. Means were based on the initial 60 secs. of each of the 10 reaction periods.
3. Total raw increase in heart rate from base level to the highest level attained.

PGR:

1. Number of PGR responses during the total record. A response was defined as a change in resistance of 400 ohms or more.

---

[4] While a study of the discordant *S*s is certainly of interest, such an investigation was postponed due to the small number of *S*s in the discordant cells. Further work on this problem is in progress.

[5] The autonomic lability score, developed by Lacey (4), evaluates stress levels corrected for differences in base levels.

2. Mean of the two largest responses occurring during reaction periods two through ten.[6]

3. Total increase in conductance over the entire record.

Respiration:

1. Increase in respiration rate from the base level to the highest level attained.

2. The variability of mean respiration rate based on the initial 60 sec. of each of the ten reaction periods.

3. Change in the standard deviation of inspiration size from reaction period one to reaction period ten.

Temperature:

1. Magnitude of the largest *increase* in temperature regardless of the reaction period in which it occurred. Measured from the base reading of the reaction period.

2. Magnitude of largest *decrease* in temperature, similar to 1.

3. Range over the entire record.

4. Autonomic lability score of the final temperature reading, using the base period as the initial level.

Blood volume:

1. The three largest oximeter responses over the total record.

2. The number of responses over the total record.

3. The variability of response frequency based on ten reaction periods.

Table 2 shows the intercorrelations among these various measures within channels. They indicate that the heart rate, PGR, and temperature measures have the highest intrachannel consistency, while respiration and blood volume measures show much less communality. In general, these correlations confirm the statement that no one measure derived from any autonomic channel is likely to be representative of general reactivity in that channel.

In order to test the major hypothesis concerning differences in autonomic reactivity between high and low perceivers, an over-all rank order was established for each channel. This order was achieved by summing the ranks on each constituent measure and then establishing a new rank order based on these sums. Similarly, a total reactivity rank order was established, based on all five channels. Table 3 presents these ranks for the two groups. The rank difference between the two groups is significant for three channels (heart rate, PGR, and respiration) as well as for the over-all total reactivity score. As expected, the latter shows the best differentiation between the two groups. It should be pointed out that no significant differences were found between the two groups in base levels on any autonomic channel. Thus, *S*s who consistently *report* a high autonomic reactivity also show high autonomic reactivity when compared with *S*s who report little autonomic feedback.

TABLE 2

INTERCORRELATIONS AMONG MEASURES WITHIN CHANNELS

Heart rate:	1 vs. 2	.75	Respiration:	1 vs. 2	.49
	1 vs. 3	.84		1 vs. 3	.03
	2 vs. 3	.93		2 vs. 3	.43
PGR:	1 vs. 2	.27	Blood volume:	1 vs. 2	.52
	1 vs. 3	.71		1 vs. 3	.14
	2 vs. 3	.47		2 vs. 3	−.00
Temperature:	1 + 2 vs. 3	.58			
	1 + 2 vs. 4	.32			
	3 vs. 4	.47			

Table 3 also indicates the large variability in ordering from channel to channel. This is consistent with Lacey's previous finding of autonomic response specificity (5). We do not know, however, whether perceptual specificity is related to autonomic specificity, i.e., whether differential report of a particular channel is also related to differential autonomic behavior within that channel. Such a comparison is not possible when extreme *S*s are selected on the basis of *total* APQ scores.

Initially, the question was raised whether *S*s who report a high level of autonomic activity differ in intensity of autonomic response or whether they over-react to autonomic stimulation. Up to this point, the results tend to favor the former interpretation. However, it is possible to evaluate the second hypothesis more directly from the data.

As was indicated earlier, the APQ contains questions relating to all five channels tested in this study. On the basis of their scores on these questions, *S*s can be assigned ranks on the perception of these channels similar to the ranks presented in Table 3. Whenever an *S*'s rank on the *perception* of a particular autonomic channel is higher than his rank on autonomic *response*, he may be said to be overestimating his autonomic reactivity; when these ranks are reversed, he may be thought of as underestimating his reactivity. By this procedure, *S*s can be classified as to their general tendency, i.e., in all five channels, to over- or to underestimate their autonomic activity. Twenty-two of the 23 *S*s could be classified in this manner. Table 4 shows the distribution

[6] The first reaction period was eliminated since it always included the largest drop in resistance, primarily as a function of shifting *S* from the resting period to the task situation.

TABLE 3
SUBJECTS' RANKS ON FIVE CHANNELS AND TOTAL REACTIVITY

Subjects	Variables					Total Reactivity
	Heart rate	PGR	Temperature	Respiration	Blood Volume	
High perceivers:						
1	9	19	7.5	11	—	13.5
2	5	1	15	—	4	3
3	3	10	2	2	—	1
4	16	14	19	3	3	10.5
5	1	15	16	10	5.5	8
6	21	5	13	5	17	16.5
7	2	9	1	13	2	2
8	—	6	11.5	4	—	5
9	13	4	7.5	12	9	7
10	4	12	21	7	13	12
11	8	—	17	8	11	10.5
12	12	7	5.5	14.5	—	9
13	15	11	4	16	15	16.5
14	6	2	14	6	5.5	4
Low perceivers:						
1	10	20	10	18	1	15
2	17.5	21	5.5	22	14	20.5
3	11	3	3	17	7	6
4	17.5	17	20	1	19	19
5	20	13	11.5	9	18	18
6	19	22	23	14.5	10	22
7	14	8	9	19	8	13.5
8	7	18	18	21	16	20.5
9	22	16	22	20	12	23
	Significance of difference between groups					
(White's T test)	$p < .01$	$p < .01$	$p > .05$	$p < .05$	$p > .05$	$p < .01$

Note.—Missing ranks indicate non-usable polygraph records. In these cases the total ranks are based on prorated estimates.

TABLE 4
AUTONOMIC PERCEPTION AND ESTIMATION TENDENCIES

	Number of *S*s Showing Discrepancies between Perception and Reactivity	
	Perceptual Overestimation	Perceptual Underestimation
High Perceivers	9	4
Low Perceivers	1	8

of these *S*s among the high and low APQ groups. The significant chi square ($p < .01$) indicates that high perceivers tend to overestimate their reactivity while low perceivers tend to underestimate. In relation to the initial question, it appears that both processes appear to be operating. *S*s who report high autonomic feedback not only show a high degree of autonomic reaction but also tend to overestimate that stimulation; similarly, low perceivers show less activity and also underestimate it.

Further studies are in progress to answer some of the questions arising from this investigation:

1. What is the relationship between autonomic feedback and autonomic reactivity for an unselected population, i.e., when dealing with the full range of perceptual and autonomic scores rather than with selected extremes.
2. What is the relationship between task performance and autonomic feedback? In the present study, the tasks were deliberately constructed to be of such great difficulty that no useful individual differences in performance could be observed.
3. What is the relationship between this perceptual-autonomic aspect of the anxiety phenomena and other behavioral indices?
4. Can response specificity be demonstrated in both the perceptual and the autonomic systems?

SUMMARY

Two groups of *S*s were selected who showed consistency in their report of autonomic activity (autonomic feedback). High perceivers were *S*s who reported a high level of autonomic feedback in anxiety situations in general and also reported such a high level in a specific stress situation. Low perceivers were *S*s who reported a low degree of autonomic feedback in both situations. The results showed:

1. Positive correlations between scores on an autonomic perception questionnaire and other paper-and-pencil tests of anxiety reactions.
2. High perceivers showed significantly greater autonomic reactivity than low perceivers.
3. High perceivers tended to overestimate their autonomic responses, while low perceivers tended to underestimate them.

REFERENCES

1. BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.
2. CAMERON, N. *The psychology of behavior disorders: A bio-social interpretation.* Boston: Houghton Mifflin, 1947.
3. JAMES, W. *The principles of psychology.* Vol. II. New York: Henry Holt, 1890.

4. LACEY, J. I. The evaluation of autonomic responses: Toward a general solution. *Annals N. Y. Acad. Sci.*, 1956, **67,** 123–164.
5. LACEY, J. I., BATEMAN, D. E., & VAN LEHN, R. Autonomic response specificity: An experimental study. *Psychosom. Med.*, 1953, **15,** 8–21.
6. LANGE, C. *The emotions.* Denmark, 1885. (Trans. by H. Kurella) Leipzig: Theodor Thomas, 1887.
7. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47,** 166–173.
8. MAY, R. *The meaning of anxiety.* New York: Ronald Press, 1950.
9. SKINNER, B. F. The operational analysis of psychological terms. *Psychol. Rev.*, 1945, **52,** 270–277.
10. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.
11. WENGER, M. A. Emotion as visceral action: An extension of Lange's theory. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950.

Received June 6, 1957.

INTRUSION OF ASSOCIATIVE RESPONSES INTO SCHIZOPHRENIC CONCEPTUAL PERFORMANCE[1, 2]

LOREN J. CHAPMAN
University of Chicago

Two traditionally opposed but not necessarily mutually exclusive explanations have been advanced to explain schizophrenic thinking disorder and, particularly, to explain the concept-forming disturbance. The first explanation is that there is a primary loss of ability. The second is that there is what might be called an increased distractibility, i.e., an inappropriate utilization of distracting material which is connected in various ways to the matter at hand. This first explanation is expounded most clearly in the writings of those who view the thinking disorder in terms of the "concrete-abstract" dimension. Prominent in this school is Goldstein (14), who conceives the loss of the "abstract attitude" as the cause not only of the lowered ability to perform on his conceptual tests but also of a wide variety of other thinking aberrations. In his view, inappropriate responses occur by default when the patient does not have available a correct conceptual response.

The second point of view is that the schizophrenic has not lost ability per se, but is prevented from demonstrating his ability by an over-responsiveness to distracting material. Cameron (4, 5) has most vigorously espoused this point of view in his controversy with Hanfmann and Kasanin (15).

Three previous studies (8, 9, 10) have demonstrated that over-responsiveness to various kinds of distracters accounts in part for the lowered performance of schizophrenics. It has been shown that a *differential* over-responsiveness to certain kinds of stimuli rather than others may account for some clinically observable features of the thinking disorder. In each of these studies, there was introduced, as the independent variable, degrees of opportunity to engage in errors of specific kinds hypothesized to be important in schizophrenia. In each case, it was found that on items in which the opportunity to participate in the hypothesized sort of deviant thinking was reduced, the schizophrenics more nearly approximated the performance of a normal control group.

The present study deals with another feature of schizophrenic thinking disorder, inappropriate associative responses. Schizophrenics often appear on clinical observation to be associating to stimuli rather than responding in a more goal-directed manner, as has been noted by well-known psychopathologists (2, 17).

This study offers a task in which errors consisting of response to associative connections may be distinguished from errors of other sorts. On such a task, schizophrenics are expected to make more such associative errors than do normals. In addition, the study tests, for the special case of associative errors, the two traditionally opposed explanations of schizophrenic thinking disorder described above.

The rationale is as follows: If susceptibility to associative distraction is the primary defect, then as the associative strength of distracters increases, the error difference between the groups should increase, at least up to a point. These increased errors should be predominantly associative in nature. On the other hand, if lowered ability to form concepts is the cause of the inferior conceptual performance and of the associative sortings, then the difference between normal and schizophrenic groups on these two aspects of performance should, at least up to a point, increase with a rise in the difficulty of concepts.

METHOD

S was required to sort a card, called here a sorting card, bearing the name of a common object, with one of three guide-card words, choosing the one of the three guide-card words belonging to the same conceptual class as the sorting card word. Each set of three guide-card words consisted of one that shared a concept with

---

[1] The author is indebted to Julian H. Pathman and George Faibish of the Downey VA Hospital for assistance in obtaining the schizophrenic *Ss*, and to Louis Halperin and Meyer Williams of the West Side VA Hospital in Chicago for assistance in obtaining normal *Ss*.

[2] The research on which this article is based was done while the author was a postdoctoral research fellow of the National Institute of Mental Health.

the sorting-card word, one which had an associative connection to it, and one which was irrelevant. For example, *S* was required to sort conceptually the word "horse," choosing from among "mule" (conceptual word), "shoe" (associative distracter), and "bottle" (irrelevant word).

Three different bases of associative connection between the sorting card word and the distracter word on the guide card were used:

1. contiguity of objects: e.g., hat-head; raincoat-water
2. contiguity of words: e.g., horse-shoe; gold-fish
3. rhyme: e.g., plate-gate; dancer-cancer

Each sorting-card word was used on various occasions with four different combinations of three guide-card words. The combinations were taken from sets of five guide-card words, such as those shown in Table 1. These four combinations defined the four experimental conditions and were as follows:

1. hard-concept word, high-associative distracter word, and irrelevant word
2. hard-concept, low-associative distracter, and irrelevant
3. easy-concept, high-associative distracter, and irrelevant
4. easy-concept, low-associative distracter, and irrelevant

Since each of these four conditions occurred with the three different bases of associative connection, there were 12 subconditions in all. Thirty items were constructed for each of these subconditions, making a total of 360 items.

Preliminary Development of Test Items

For establishing differences in conceptual difficulty, concepts were offered to judges two pairs at a time (e.g., horse, mule; horse, chicken), and the judges were asked to decide which pair formed the best concept, goodness of concept being defined as the degree to which the two named things are alike in some way. The pair designed to form a relatively poor concept was intended to have an identifiable concept but one less obvious than the pair designed to have a "good" concept.

To establish differences in associative strength, two pairs of words were presented at a time (e.g., horse, shoe; horse, cream), and the judges were asked to decide which pair had the higher associative connection between them, defined as the degree to which one word calls to mind the other. *E* chose the low association words by the same criterion as the irrelevant, i.e., they were intended to have minimal associative connection with the sorting-card words.

The judges were groups of psychology students (28 to 97 per group). On the goodness-of-concept judgments, the median agreement was 98 per cent for the 30 items in each of the three bases of association. In the associative strength judgments, it was necessary to accept differing standards of agreement among the different bases of association. The median percentage agreements for the word pairs were 100 for contiguity of objects, 98.5 for contiguity of words, and 92.5 for rhymes. The high association, low association, and irrelevant categories were roughly equated for frequency of occurrence in the language according to Thorndike and Lorge's (20) word counts.

TABLE 1

SAMPLES OF WORDS USED IN ITEMS

Sorting-Card Words	Guide-Card Words				
	High Asso-ciation	Low Asso-ciation	Irrelevant	Easy Concept	Hard Concept
hat	head	street	garage	cap	dress
raincoat	water	food	pencil	umbrella	awning
lip	stick	shoe	chair	tooth	knee
gold	fish	ache	typewriter	platinum	steel
dragon	wagon	handle	rope	sea serpent	trout
face	lace	wagon	weed	toe	liver

Subjects. One normal and one schizophrenic group, each consisting of 32 adult males, were tested. The normal *S*s were medical and surgical patients at the West Side VA Hospital, Chicago, Illinois, for whom there was no indication either in the hospital record or in an interview of psychiatric or neurological disorder. The schizophrenic group, patients in the Downey VA Hospital, had all received a hospital diagnosis of chronic schizophrenia, with their first hospitalization for psychosis occurring at least one year and their last at least 3 months previous to testing. Other prerequisites for selection were absence of any indication of either cerebral organic involvement or mental deficiency, no current drug therapy, no electric shock therapy within the past 3 months, nor more than 35 shock treatments altogether. In addition to the *S*s in the schizophrenic groups, 16 schizophrenic *S*s were also seen but were not included in the final group because they were unwilling or unable to participate.

Mean age was 38.7 years for the schizophrenics and 37.0 for the normals, while mean years of education was 10.7 for the schizophrenics and 10.1 for the normals. Although the groups did not differ significantly with respect to age or education, the schizophrenic group was significantly lower on functioning IQ as determined by the CVS (16), presumably due to the disruptive effects of schizophrenic pathology. On this test, the mean IQ was 108 for the normals, 90 for the schizophrenics.

The *S*s were assumed to be of approximately equal premorbid intelligence because of the equivalent education. An additional approximation of premorbid intelligence was obtained from the item number of the last vocabulary word answered correctly, a measure introduced by Becker (3). This appears to be a better estimate than the total number of vocabulary words answered correctly because one effect of schizophrenia is to produce an unevenness of performance. On this measure, the schizophrenics scored a mean of 8.6, the normals, 9.4. The difference was not significant ($p = .22$).

Although no selection was attempted by diagnostic subtype, the sample was found to include 18 patients diagnosed as paranoid, 9 as hebephrenic, and 5 without classic subtype.

Apparatus. A 13″ x 10″ x 4″ box was used for sorting the cards, the box having three 3¼″ slots along the front edge of the top through which the sorting cards could be deposited. A wire rack for holding the guide cards was placed on top of the box in back of the slots.

The three guide-card words were typed on a 2″ x 11″ card, and were spaced so that when the guide card was placed in the wire rack, each of the three words was in back of one slot. The sorting cards were 3″ x 5″.

Experimental procedure. Before the testing proper, *E* presented six practice items to make certain that *S* understood the nature of the task. These items were identical in form to those in the high association condition of the experiment proper. The only thing that differed was that *E* asked *S* to explain his sorting, and *E* helped him with them. The data from these items were not included in the analyses of results.

After the six practice items, *E* presented 45 items from one of the four main experimental conditions. *E* placed the bundle of 45 guide cards in the rack on top of the box and handed *S* the pack of 45 sorting cards. *E* said, "Now here's a bunch more to sort. For each one, you decide which word up here mentions something that can be called by the same name as the thing mentioned on the card in your hand, and you drop your card in the correct slot, but this time you needn't tell me about it." At the beginning of presentation of the cards for each of the other three conditions *E* again said, "You decide which thing mentioned up here can be called by the same name as the thing mentioned on the card in your hand."

It will be recalled that each sorting card was used with four different guide-card items. Therefore, in order to minimize resultant transfer effects, each *S* responded to only 180 of the 360 sorting card items (15 items in each of the 12 subconditions), receiving the same sorting card in the high-association—hard-concept items as in the comparable low-association—easy-concept items, and similarly receiving the other half of the sorting cards both with the low-association—hard-concept items and with the comparable high-association—easy-concept items. The blocks of cards in each of the four main experimental conditions were presented in a counterbalanced order among *S*s. Within each of the four conditions, items based on the three different bases of association were intermixed in a prearranged random order that was the same for all four main conditions and was constant among *S*s. On the guide cards, words bearing the three kinds of relationship to the sorting word were randomized as to position.

RESULTS

Table 2 shows the total error scores under the four main experimental conditions. In evaluating all the differences, nonparametric techniques were used because the data were extremely skewed.

TABLE 2
MEAN TOTAL ERROR SCORES

Condition	Normals	Schizophrenics
Low Association		
Easy concepts	.09	1.62
Hard concepts	.81	3.38
High Association		
Easy concepts	.66	4.03
Hard concepts	3.09	11.41

As expected, the schizophrenics made significantly more errors than normals ($p < .01$) in all four conditions. This and all other between-group differences were evaluated by a two-tailed Mann-Whitney *U* test (22).

In addition, manipulation of each of the two independent variables produced significant error differences for each group. With associative strength held constant, both schizophrenics and normals made significantly more errors when concepts became harder, whether associative strength was high or low, and similarly, each group made significantly more errors when the associative strength of the distracters became stronger, both for easy concepts and for hard ($p < .01$ in each case). These and all other within-group differences were evaluated by a two-tailed Wilcoxon test (22).

Of central interest, however, is the amount of divergence of the error curves for the two groups under the various conditions. While this study was primarily concerned with whether the two independent variables are related to *difference* in frequency of errors between schizophrenic and normal *S*s, the effects of the two independent variables cannot meaningfully be compared with one another. However, whether each produces a significant change in errors can be determined.

First, consider the concept difficulty dimension. When associative distracter strength was held low and concepts were varied from easy to hard, the errors of the schizophrenics rose more than those of the normals but not significantly so ($p = .16$). However, when the associative strength was held high, varying the conceptual difficulty in the same way produced a significantly greater rise in errors of schizophrenics than of normals ($p < .01$). Also, a direct comparison of the two divergences revealed that the concept difficulty dimension produced a significantly greater effect in the high association condition than in the low ($p < .01$). Thus, one may conclude that difficulty of concepts increases the inferiority of schizophrenics to normals, but the inferiority is intensified when the test material contains opportunity for incorrect sorting by strong associative distracters.

Next, consider the effects of manipulating strength of associative distracter with difficulty of concepts held constant. For easy concepts, there was a greater rise in errors for schizophrenics than for normals as associative strength increased ($p < .02$). However, for hard concepts, the effect of manipulating associative strength was much greater. This divergence was significant at less than the .01 level, and the difference between the two divergences was similarly significant. Thus, it may be concluded that as the associative strength of the distracters increases, the schizophrenic decrement increases, but the effect of the associative strength dimension is much greater when the correct conceptual solution is more difficult.

Now compare distracter placement errors with irrelevant placement errors for those two conditions under which the associative strength of distracters was held high. Table 3 shows these comparisons, listing the errors separately for the three bases of association.

TABLE 3

MEAN DISTRACTER AND IRRELEVANT ERROR SCORES OF NORMAL AND SCHIZOPHRENIC *S*s, SORTING ON THREE BASES OF HIGH ASSOCIATIVE DISTRACTION

Association Basis	Normals		Schizophrenics	
	Dist.	Irr.	Dist.	Irr.
Contiguity of objects	2.03	.09	6.75	.31
Contiguity of words	1.03	.09	5.03	.62
Rhyme	.38	.12	1.91	.81

First, it should be noted that most of the errors consisted of placements with the distracter word, accounting for 89 per cent of the schizophrenics' errors and 92 per cent of the normals' errors in the high association condition.

Errors other than in response to the associative distracter should, by chance, be equally divided among placements with the two incorrect guide-card words. Therefore, to measure errors arising from response to the associative distracter, the number of distracter placements for each *S* was corrected by subtracting from it the number of irrelevant placements. This score of distracter placements minus irrelevant placements should yield a comparatively pure measure of response to the associative connections. Table 4 shows these difference scores for each condition. Because, as noted earlier, the low association and irrelevant words had a similarly low associative connection to the sorting word, distracter-minus-irrelevant scores should approximate zero in the low association condition. Table 4 tends to bear out this prediction.

The analyses which were done for total error score were recomputed, using this distracter-minus-irrelevant score. The results of these analyses were similar to the results reported above for total error scores. On this difference score, each group increased significantly in response to increased difficulty of concepts ($p < .01$ in each case). This means that not only does each group make more total errors when concepts become more difficult, but also, each makes more errors specifically resulting from response to associative distracters. Too, a between-group comparison showed that the schizophrenics significantly exceeded the normals in the extent to which this score rose with an increase in difficulty of concepts ($p < .001$). Thus, as concepts get harder, not only do schizophrenics make more errors than normals; they increase more than normals in identifiable associative errors. In addition, as with total error score, the distracter-minus-irrelevant error score of the schizophrenics diverged more from that of the normals in response to the difficulty dimension when high associative distracters were present than when low ones were present ($p < .01$).

Similar analyses were done to determine the effect of the associative strength variable on distracter-minus-irrelevant error scores. As associative strength of distracters rose, again both groups made higher scores on this difference score just as they did on total errors ($p < .001$ in each case). Again, the schizophrenics exceeded the normals in the extent to which this error score increased as a function of the associative strength of distracters ($p < .001$). Also, the error curve of the schizophrenics again diverged more from that of the normals in response to the associative strength dimension when concepts were hard than when they were easy ($p < .01$).

TABLE 4

MEAN DISTRACTER-MINUS-IRRELEVANT SCORES

Condition	Normals	Schizophrenics
Low Association		
Easy concepts	−.06	.12
Hard concepts	−.12	.21
High Association		
Easy concepts	.59	3.53
Hard concepts	2.53	8.41

In conclusion, both independent variables had a strong effect on associative responses, and the two strongly interacted. It is clear, then, that both difficulty in forming concepts and susceptibility to associative distraction are in part responsible for the fact that schizophrenics show more associative intrusions than do normals. Some substantiation is found here for both hypotheses.

DISCUSSION

A supplementary inference from these data is that some aspects of the schizophrenic thinking deviancy are exacerbations of normal response tendencies. The associative distracters, responded to by the schizophrenics, were invented by normals and were judged as having associative connection by other normals. The errors of the normal *S*s, though fewer in number, were, like those of the schizophrenics, predominantly associative rather than irrelevant placements. Also, the normals, like the schizophrenics, increased their associative choices in response to each of the two independent variables. Thus, the schizophrenics differed from the normals on this task primarily on the frequency of these responses.

Similarly, in each of three previous studies (8, 9, 10), the specific kind of error which accounted primarily for the difference between schizophrenics and normals tended consistently to be committed more often by normals than other types of errors which had an equal opportunity to occur by chance.

It has been suggested by various writers that the thinking errors of schizophrenics resemble those made by normals under this or that special condition. Among the special conditions mentioned have been sleep (13, 17), relaxed attention as in free association (17, p. 17), drugs (18), and fatigue (5, 12).

Cameron and Magaret (6) also reported a finding that normals in completing incomplete sentences in an ordinary classroom situation showed kinds of errors that seemed to resemble schizophrenic thinking disorder. However, when they discuss in detail elsewhere (7) the broader issue of the many similarities of normal errors to those of schizophrenics, they emphasize the importance of special conditions in producing such effects in normals. Among their special conditions are the interruption of ongoing activity, environmental changes, preoccupation, and emotional excitement.

The present findings suggest that the error patterns of normals under ordinary functioning are found exacerbated in what is called schizophrenic thinking disorder. There is supporting evidence for this hypothesis in the data of other investigators. Fey (11) found that on the Wisconsin Card Sorting Test, the perseverative errors most often made by normals were also the kind primarily responsible for the fact that the schizophrenics erred more than the normals. This is especially impressive, considering that the test is constructed so that if responses occurred in a random fashion the perseverative errors would be a little less than half as frequent as the nonperseverative errors.

Further evidence for this kind of relationship is found in the apparent similarity between other solution preference patterns or "biases" of normals, and well-known features of schizophrenic thinking pathology. An example is seen in errors in syllogistic reasoning. Although Von Domarus (21) and Arieti (2) state that a distinctive feature of schizophrenic thinking is a tendency to conclude that two subjects are identical on the basis of their sharing common qualities, Sells (19) and Wilkins (23) have reported that the same error tendency is strong among normals.

Another well-known schizophrenic error pattern is the interrelating or combining of ideas in an inappropriate way. This is especially prominent in delusional patients. However, Abel (1) reported that when her normal *S*s attempted to reproduce a difficult passage of scientific prose, about half of them produced sentences in which two or more ideas were inappropriately combined into a whole so that the original meaning was lost.

These latter two similarities between normal and schizophrenic thinking are only suggestive because they have not been measured on the two populations by the same tasks. An important direction for new research would seem to be the designing of tasks measuring different types of error preferences to determine whether there are valid correspondences between normal and schizophrenic errors. Such relationships might be sought for facets of the schizophrenic thinking disorder not readily referable to already isolated normal biases. Also, one might determine the extent in schizophrenia of previously isolated normal biases which are not obviously present in clinical schizophrenic psychopathology.

Clinically, schizophrenics appear to differ greatly among themselves in their thinking disorder. It would be valuable to determine whether such differences may be described as exacerbations of different normal biases, or whether some characteristics of their thinking disorders are not found in normal *S*s.

SUMMARY

The present study tested the hypothesis that schizophrenics' substitution of associative responses for correct responses is due to heightened susceptibility to associative connections, and also tested the alternative hypothesis that it is due to a primary loss of ability to perform correctly. A conceptual card-sorting task was used in which for each sorting choice there were three alternatives, one of which was the correct conceptual response, one an associative distracter, and one neither. The difficulty of the correct conceptual task was varied at two levels, testing the hypothesis that difficulty in forming concepts is responsible for the schizophrenics' greater number of errors, in particular their associative errors. Similarly, there were two levels of the associative strength of the distracter, testing the hypothesis that susceptibility to associative distraction, independent of difficulty of the correct task, is a source of these errors.

Both hypotheses were supported in that both stimulus variables were found related to frequency of associative errors for both schizophrenics and normals, with the divergence of the error scores of the two groups resulting from a greater susceptibility of the schizophrenics to each of the two independent variables. The implications of these results were discussed in terms of the suggestion that some aspects of schizophrenic thinking pathology consist of exacerbations of normal response tendencies.

REFERENCES

1. ABEL, THEODORA M. Unsynthetic modes of thinking among adults: A discussion of Piaget's concepts. *Amer. J. Psychol.*, 1932, **44**, 123–132.
2. ARIETI, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.
3. BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.
4. CAMERON, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
5. CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. II. New York: Ronald, 1944.
6. CAMERON, N., & MAGARET, ANN. Experimental studies of thinking. I. Scattered speech in the responses of normal subjects to incomplete sentences. *J. exp. Psychol.*, 1949, **39**, 617–627.
7. CAMERON, N., & MAGARET, ANN. *Behavior pathology.* Cambridge, Mass.: Riverside, 1951.
8. CHAPMAN, L. J. Distractibility in the conceptual performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1956, **53**, 286–291.
9. CHAPMAN, L. J. The role of type of distracter in the "concrete" conceptual performance of schizophrenics. *J. Pers.*, 1956, **25**, 130–141.
10. CHAPMAN, L. J., & TAYLOR, JANET A. The breadth of deviate concepts used by schizophrenics. *J. abnorm. soc. Psychol.*, 1957, **54**, 118–123.
11. FEY, ELIZABETH T. The performance of young schizophrenics and young normals on the Wisconsin Card Sorting Test. *J. consult. Psychol.*, 1951, **15**, 311–319.
12. FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
13. FREUD, S. The unconscious. *Collected papers.* Vol. IV. London: Hogarth Press, 1953.
14. GOLDSTEIN, K., & SHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
15. HANFMANN, EUGENIA, & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, No. 67.
16. HUNT, W. A., FRENCH, ELIZABETH G., KLEBANOFF, S. G., MENSH, I. N., & WILLIAMS, M. Clinical possibilities of an abbreviated individual intelligence scale. *J. consult. Psychol.*, 1948, **12**, 171–173.
17. JUNG, C. G. The psychology of dementia praecox. *Nerv. ment. Dis. Monogr. Ser.*, 1936, No. 3.
18. RINKEL, M., HYDE, R. W., & SOLOMON, H. C. Experimental psychiatry. III. A chemical concept of psychoses. *Dis. nerv. System*, 1954, **15**, 259–264.
19. SELLS, S. B. The atmosphere effect: An experimental study of reasoning. *Arch. Psychol.*, 1936, No. 200.
20. THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words.* New York: Columbia Univer., 1944.
21. VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. Calif. Press, 1944.
22. WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Henry Holt, 1953.
23. WILKINS, M. C. The effect of changed material on ability to do formal syllogistic reasoning. *Arch. Psychol.*, 1928, No. 102.

Received June 7, 1957.

THE RELATIONSHIP BETWEEN JUDGMENTS OF A COMMUNICATOR'S POSITION AND EXTENT OF OPINION CHANGE[1]

WALTER WEISS

Boston University

IN SOME situations, persons to be exposed to a communication may entertain prior expectations concerning the communicators' position that can influence their interpretation of the communication, or its effectiveness. Such anticipatory categorization may result when the audience has had past contact with the communicator or with a similiar class of communicators. But in many circumstances, it is on the basis of the content of the communication itself that the audience comes to view the source as holding a particular opinion position. This judgment, whether vaguely or precisely formed, is presumed to be related to the extent of opinion change exhibited by members of the audience. In the extreme case, misinterpretation of the sense of the communicator's arguments may lead to no change or to a negative effect. But when the general direction of the communicated opinion position is obvious and is opposed to that held by the audience, is there a relationship between the extremeness of the opinions attributed to the communicator and extent of opinion change?

There is some evidence of a positive relationship between extent of opinion change and extent of discrepancy between the communicator's position and that of the initial position of the recipient (1, 3, 9). But as Hovland and Pritzker (4) point out, this naturalistic association may be due in part to a "ceiling effect" (i.e., those whose initial opinions are least discrepant from the communicator's position can change least) or to differential regression toward the mean on the retest. In their own study, amount of change advocated was made independent of the *S*s' initial positions and the phenomenon was confirmed. However, an elaborated, persuasive communication of the kind usual in a "natural" setting was not employed. Accordingly, judgment of the communicator's position was not a variable in their research.

PROCEDURE

An after-only arrangement was employed, involving three groups: two exposed to a communication on juvenile delinquency and a control group that was not. One experimental group read an appeal by an *anonymous* communicator for the adoption of a decidedly, though not extremely, punitive policy toward delinquents. Arguments supporting this position were essentially paraphrases with elaborations and suitable changes of four statements from Form B of the Wang-Thurstone scale of attitude toward the treatment of the criminal. These statements ranged closely around a median value of 8 (11 being the extreme punitive end of the scale). The second experimental group read the same basic communication but with the addition of a few introductory statements to the effect that some people hold to the opinion extremes of complete leniency or excessive punitiveness toward delinquents. An opinion statement was given illustrating each of these end anchors for the opinion continuum.[2] Then the communicator presented his own position, which was identical in content to the material read by the other experimental group.

After reading the two-page article, the *S*s responded to a modification of Form A of the Thurstone scale *as though they were the communicator*; that is, they signified agreement or disagreement with each of the 20 statements as they thought the communicator would, now that they were aware of his position. Following this, they answered a modification of Form B of the scale in terms of their *own* opinions on the problem of combatting delinquency. A series of general judgmental items completed the questionnaire.

The no-communication control group rated

---

[1] The research reported in this article was supported initially by funds from the Office of Naval Research, under Contract NONR-492(04), and completed under grant M-1351 from the National Institute of Mental Health.

[2] The expectation was that the extreme opinion anchors would induce a contrast effect (6, 9) or extension of the scale of judgment (7, 8), with a consequent decrease in the perceived extremeness of the communicator's position. In answer to a direct question, over 80 per cent of the students in each group did consider the opinion statements as representative of the extremes of the opinion continuum.

each of the statements of Form A on a 1–11 basis in terms of its expressed degree of approval or disapproval concerning the use of punishment. Then members of this group gave their own opinions on delinquency by responding to Form B. They also answered a few judgmental items.

Booklets covering the three treatments were arranged in a random sequence, following a table of random numbers. They were handed out individually to the *S*s who were seen during regular class periods at Boston University and American International College.[3] The purpose of the study was stated to be an examination of people's judgments of the opinions of others on a controversial topic.

RESULTS

The Wang-Thurstone scales were scored by assigning a value of one to a response favoring punitiveness toward delinquents and a zero to a response opposing punitiveness. Neutral items on each scale could not be used under this system. Accordingly, data on the judged position of the communicator are based on responses to 18 items of Form A, and data on the *S*s opinions are based on 17 items of Form B.[4]

When judgments of the communicator's opinion position are examined, it is found that the group that is provided the end anchors locates him at a significantly *less* punitive position than does the other communication group (10.67 vs. 13.30; $p < .001$[5]). This result is supported by congruent and significant differences between the two groups in their responses to two specific questions concerning the author's views on the effectiveness of punishment and guidance in treating delinquents. Thus, judgments of the communicator's position were affected in the desired manner despite identity of the basic facts and arguments supporting and elucidating his viewpoint.[6]

TABLE 1
MEAN AFTER-OPINIONS ON DELINQUENCY

	Communication Groups		
	No standards	End standards	Control group
Mean	6.18	5.22	4.29
SD	2.77	2.53	1.70
N	60	59	68

Analysis by *t* tests of the opinion data presented in Table 1 indicates that the communication group *not* provided the end standards exhibits a significantly *more* punitive after-mean than does the other experimental group ($p = .05$). Thus, it is the group exhibiting the more extreme judgment of the communicator's opinion position that shows the greater effect of the communication. The after-mean of each experimental group is significantly more punitive than that of the control group ($p <$ .03 and $< .001$).

The difference between the judged position of the communicator and the opinion mean of the control group can be conceived of as an indication of the amount of change advocated. Thus, for the group provided the end standards, the amount is 6.38 and for the other group 9.01. Using the control mean as the baseline, the amount of change exhibited is 0.93 and 1.39 respectively. If average amount of change is divided by amount advocated, an index of relative change can be obtained. For the group without the end standards, this is 21.0 per cent of the amount advocated and for the other group only 14.6 per cent. (The same result is obtained when the Wang-Thurstone scale values are used.) These data, though crude, do not agree with the finding of Hovland and Pritzker (4) of a negative relationship between relative change and amount of change advocated and obtained.

Data concerning the relationship between judgments of the communicator's position and the *S*'s after-opinions on delinquency are presented in Table 2. By analysis of variance, the main effect of judgment of author's position and the interaction effect were not sig-

---

[3] The author wishes to thank Malcolm Klein, of Boston University, and Richard Sprinthall, of American International College, for distributing the booklets to their respective classes.

[4] None of the results is changed if the items are scored according to the original scale values given by Wang and Thurstone for attitude toward the treatment of the *criminal*.

[5] All p values are two-tailed.

[6] When the *S*'s judgments of the author's opinion position are scored in terms of the original Wang-Thurstone scale values, the median for the group given the end standards is 6.4 and that for the other group is 7.3. Since the four statements embodying the communicator's position had a median value of 8.0, the group that was not provided the end standards seems to have judged its position more "accurately" than did the other group.

TABLE 2

RELATIONSHIP BETWEEN *Ss*' OPINIONS ON DELINQUENCY AND THEIR JUDGMENTS OF AUTHOR'S POSITION

Author's position	*Ss*' Opinions	
	No standards	End standards
Above mdn.	6.51 (41)	4.86 (21)
Below mdn.	5.47 (19)	5.42 (38)

TABLE 3

RELATIONSHIP BETWEEN AFTER-OPINIONS ON DELINQUENCY AND JUDGMENTS OF COMMUNICATOR'S TRUSTWORTHINESS

	No standards	End standards
Relatively trustworthy	6.86 (22)	5.45 (38)
Other	5.79 (38)	4.75 (20)

nificant. Also, the judgment subgroups within each experimental group were not significantly different from each other. It is only among those who locate the communicator above (i.e., more punitive than) the median judgment of the total group that the communications have significantly different effects (t test p value .05).

In response to a multiple choice question, a significantly higher proportion of people given the end standards rated the communicator as relatively trustworthy than did those in the other experimental group (65.5% vs. 36.7%; $p < .01$). Analysis of the data in Table 3 by analysis of variance indicates that the main effect of judgment of trustworthiness and the interaction effect were not significant. Differences between subgroups within a communication group were also not significant.

Judgments of trustworthiness were found to be negatively related to judgments of the communicator's opinion position. That is, those who viewed the communicator as relatively trustworthy judged him to be significantly less punitive than did those who did not consider him to be trustworthy. When the judgments of communicator trustworthiness and opinion position are combined, it is only between those subgroups rating him as relatively trustworthy *and* judging his position to be above the median for all experimental *Ss* that the difference between the two treatments is statistically significant. When either judgment of trustworthiness or of opinion position is held constant, the other judgment is still not associated significantly with the after-opinions of the *Ss*.

Two specific, multiple choice questions sought the *S*'s own judgments of the effectiveness of guidance and punishment as methods for dealing with delinquents (i.e., with the instrumental value of such procedures). Within each group and for the aggregate of all *Ss*, responses to these items were significantly and positively related to the punitiveness of after-opinions. However, the two communication groups did not differ significantly in the proportions judging the instrumental value of guidance or punishment to be high or low. When *Ss* in the communication groups are separated in terms of their responses to these two questions and comparable subgroups are compared, the differences in after-opinions on delinquency are in the same direction as the over-all difference between the two experimental groups.

DISCUSSION

The finding of greater opinion change for the group attributing a more extreme position to the communicator agrees with previous research relating extent of change to discrepancy between initial opinion and advocated opinion (1, 3, 4, 9). However, this judgment dimension does not seem to be related in any simple or obvious manner to the difference in effectiveness of the two communications. This is evidenced by the lack of an over-all or within-groups association between extremeness of judgment of the communicator's opinion and punitiveness of own position, and by the finding of a significant difference in communication effectiveness only between the subgroups judging the communicator's position to be quite extreme. The methodological defect of an artificial restriction or "ceiling effect" on extent of change was avoided in this study; for the position attributed to the communicator by the less influenced group was considerably more extreme than its own mean opinion and than that of the other experimental group.

Internal analyses have not yielded a satisfactory explanation for the main opinion results. The availability of information concerning the before-opinions of the *S*s and concerning their interpretations of the significance and force of the arguments might have clarified the meaning of the data. As sheer speculation, however, it is conceivable that the method employed for affecting the perceived extremeness of the communicator's position reduced the impact of the communication; that is, mention of the extreme opinion standards may have reminded the students of their own nonpunitive position or fostered a crystallization of opinion, thereby increasing resistance to change.

Hovland and Pritzker (4) have suggested that with an anonymous source, a large discrepancy between the position advocated and that held initially by the audience may diminish regard for the source. This supposition is supported by the current data. But the further expectation that such lowered regard would be reflected in lessened communication effectiveness is not confirmed. It is possible that the condition of anonymity of the communicator changed the personal meaning and functional significance of a judgment of trustworthiness. However, even with an identified and known communicator, there is some research evidence that apparent regard for the source need not be a critical factor underlying certain judgments of a communication or its effectiveness (2, 10).

In general, more information is needed on the psychological *processes* underlying opinion change, so that the influence of particular communication variables and the functional significance of relevant judgments can be fitted within and predicted from the framework of a systematic theory.

SUMMARY

The purpose of the research was to examine the relationship between judgments of a communicator's position and extent of opinion change. Three groups were formed: two exposed to a communication on juvenile delinquency and a control group that was not. One communication group read an appeal by an anonymous author for the adoption of a decidedly, though not extremely, punitive policy toward delinquents. The other experimental group read the same basic communication with the addition of a few introductory statements reminding the audience that some people hold to the opinion extremes of complete leniency or excessive punitiveness toward delinquents. Both groups then responded to a modification of Form A of the Wang-Thurstone scale of attitude toward the treatment of the criminal as they thought the communicator would, and then answered Form B in terms of their own opinions on the problem of delinquency. A series of general judgmental items completed the questionnaire. The control group rated each statement of Form A on a 1–11 scale in terms of its expressed degree of approval or disapproval concerning the use of punishment and then gave their own opinions on delinquency by answering Form B.

The group given the extreme opinion standards judged the communicator's position to be less punitive than did the other communication group. But it was the latter group that was more influenced by the communication. However, when the data for each experimental group and for the combined aggregate of experimental *S*s were examined, judgment of communicator's position was not found to be related to own opinion on the treatment of delinquents. The mean opinion position of each experimental group was more punitive than that of the control group. Although the judgment of trustworthiness of source was negatively related to extremeness of judgment of the communicator's position, it was not significantly associated with the *S*s' opinions.

REFERENCES

1. EWING, T. N. A study of certain factors involved in changes of opinion. *J. soc. Psychol.*, 1942, **16**, 63–88.
2. FINE, B. J. Conclusion-drawing, communicator credibility, and anxiety as factors in opinion change. *J. abnorm. soc. Psychol.*, 1957, **54**, 369–374.
3. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.
4. HOVLAND, C. I., & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.
5. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
6. MCGARVEY, HULDA R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol.*, N. Y., 1943, **39**, No. 281.
7. ROGERS, S. The anchoring of absolute judgments. *Arch. Psychol.*, N. Y., 1941, **37**, No. 261.

8. SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. Unpublished manuscript. Yale Univer., 1956.

9. SIMS, V. M. Factors influencing attitudes toward TVA. *J. abnorm. soc. Psychol.*, 1938, **33**, 34–56.

10. WEISS, W. Opinion congruence with a negative source on one issue as a factor influencing agreement on another issue. *J. abnorm. soc. Psychol.*, 1957, **54**, 180–186.

Received October 10, 1957. Prior publication.

CASE REPORT

WAS LARS SEXUALLY ASSAULTED? A STUDY IN THE RELIABILITY OF WITNESSES AND OF EXPERTS[1]

ARNE TRANKELL
University of Stockholm

LARS, a five-year-old boy, lived with his parents on the outskirts of a fair-sized town. The family lived in an old house, without the modern conveniences of the newly built apartments nearby. One day, Lars was left alone in the street for half an hour while his mother and grandmother visited the local stores. When they came home, they met Lars just outside the door. He had a bundle of newspapers under his arm and was eating a bit of liquorice that had stained his lips black. When his mother asked him where he had got the papers and the candy, Lars replied that he had got the papers from a "window-cleaner man." As for the candy, he said he had bought it with a big black penny, which he had also had from the window cleaner. On hearing this disturbing tale, the two women drew the boy into the entrance, and there he told them that he had also had to suck the man's penis. They all quickly went up to the apartment, where the grandmother discovered that the boy's fly buttons were undone. This fact, too, the boy explained as a consequence of his encounter with the window cleaner.

WHAT HAPPENED? TWO VERSIONS

According to what the parents told the police, the mother and father agreed over the telephone to go into the matter no further until the police had had a chance to hear the boy. The police report states that the mother visited the police station the same evening and reported the incident. It does not emerge from the report whether the child went with her to the police. A number of the boy's words are, however, reproduced as direct speech in the police report. The most important is the following: "I had to suck his doodle, and he pushed it down so that he nearly choked me. It was so disgusting." The boy said, moreover, that the man lived in the same house as Sune Olsson.

If one follows the police report, one finds that Lars was escorted to the scene of the crime as given by the mother. He pointed out a house some blocks away from his own home. The police quickly established that there lived in this house a window cleaner of foreign nationality, possibly of Russian origin. It also turned out that the window cleaner intended to travel the following day to an unknown destination, so that it was arranged to pick him up quickly. The window cleaner, a man of 24 here called S., was arrested by the police at about five P.M. at the office of the firm that employed him. Earlier the police had arranged with commendable speed to have the child's reliability tested. The examination took place between two and four P.M. on the same day that S. was arrested, that is, some 24 hours after the event. The test consisted of a conversation between the child psychiatrist and Lars' mother to establish the background, together with a tape-recorded interview with Lars. According to a statement in the report, there was also a summary test of the child's intelligence. The whole investigation took place at the police station, where the mother and the boy had been called.

---

[1] *Editor's Note.*—Expert evaluation of children's testimony has a long tradition in European forensic psychology. In Sweden, where the case reported by Professor Trankell originates, the courts have come to rely on special experts when the evidence for a charge depends mainly on a child's testimony, as it typically does in cases of assault against minors, usually committed under circumstances that prevent anyone but the accused and the purportedly assaulted child from knowing just what has occurred. If the expert finds the child to be a reliable witness, a conviction may be obtained on the basis of the child's evidence.

The case of Lars, which has the fascination of a good detective story, may serve to acquaint American psychologists with this sort of court service with which they have little if any direct experience. As a case in abnormal and in social psychology, however, its primary interest lies in another direction. Not only does Professor Trankell provide us with new illustrations of familiar phenomena concerning the unreliability of witnesses; what is novel is the way in which his account brings to our attention the more serious problem of reliability that arises when experts disagree.

It is Professor Trankell's suggestion that the experts' divergent findings in the present case follow from divergent methods of approach. One of the experts draws his conclusions from a personality appraisal of the reliability of the child witness. Such an approach is invited by the legal formula that is used in Sweden, according to Professor Trankell, when specialists are appointed: "The Municipal Court has appointed you as a specialist in the case of ____ to give your opinion on the reliability of the __-year-old child ____." The other expert, Professor Trankell himself, arrives at very different findings from an approach that he characterizes as focussed on the social relationships and interactions of the alleged victim and criminal, attempting to assess not so much the reliability of the witness as the reliability of the crucial statements upon which the case depends. To what extent is the conclusion of each expert the result of his method? The reader may well read the case with this question in mind, remembering that the report is presented by one of the parties to the disagreement.

Besides the child psychiatrist's interview with Lars, the police report included several interviews with S., S.'s fiancée, her mother, the landlord's agent in the building, the assistant in the sweetshop where Lars bought his liquorice, and S.'s employer. A summary of the police enquiry gives the following picture:

> S. stated that he had cleaned the windows of a chemical business in the street where Lars had been left alone by his mother and grandmother. When he had finished and was going to cycle on to his next place, he had been accosted by a vagrant, who had begged him for some money. He remained standing in the street for a moment, with the ladder under his arm and the bicycle beside him, while he argued with the beggar. During this discussion a little boy came up to him and asked: "May I carry your ladder, Mister?" S. claimed to have answered that the ladder was too heavy. S. had not observed the boy's appearance closely. When S. had got rid of the vagrant he pushed the bicycle on to the sweetshop, where he had a cleaning contract. He left the bicycle outside the shop on the still unfinished pavement and leaned the ladder against the wall at the side of the shop. Then he went on another hundred yards or so to his own house. He intended to have a break for coffee in his home before cleaning the shop's windows.
>
> Lars' information when he was taken to the place agreed well with S.'s. The boy independently pointed out the place where S. had laid down the bicycle and the place where he had stood the ladder. He also indicated an entrance to S.'s house, through which he claimed to have passed on his way up to "the man's house." The staircase on which S.'s flat was situated was on the inner side of the courtyard inside the long block and could not be reached from the street except through the cellar, unless one wanted to go round the whole block, which meant a detour of several hundred yards. Interrogation of the landlord's agent established that the cellar doors were always kept locked. A little child could therefore not get through the cellar by himself.
>
> S. definitely asserted that he had gone through the cellar alone and that he had unlocked the door and locked it again behind him when he passed through. He denied having heard anyone come after him through the cellar. When he came up into his flat, which was empty during the day, he made some coffee and a few sandwiches. While he was drinking his coffee, there was a ring at the door. When he opened the door he found a little boy on the stairs, who asked if he might have some old newspapers. S. left the door ajar and went into the kitchen, where he took out a bundle of papers in the cupboard above the refrigerator where such things were kept. He came back into the hall and gave the papers to the boy. After that the boy went off. S. formally denied giving the boy any money. When asked if it was the same boy who had spoken to him earlier in the day, S. could not say for certain. He thought that it was quite possible, as he had not looked at him very closely. He had not noticed if the boy had entered the flat during the short time he was himself in the kitchen. When the boy had gone, S. finished his coffee and then went back to the shop to clean the windows there. In doing so, he followed the same way back through the cellar.
>
> From the relatively few bits of spontaneous information given by Lars in the interview with the child psychiatrist, it emerges that the boy, after visiting S.'s flat, went down into the courtyard where he met another boy. He would not give this boy's name, and when the psychiatrist asked what he had done with him, Lars answered: "I was not with him." After that, Lars went to the sweetshop round the block to spend the coin he claimed to have got from the window cleaner. When he arrived at the shop "the man" was already there cleaning windows.
>
> On the occasion of a second police visit to the scene of the crime, a number of questions were put to Lars. The answers seemed to show that the boy had a good idea of what S.'s flat looked like. When asked if he had been inside the flat, he answered that he had been in the kitchen. Asked if he had been into the bathroom, he said that he had, and when asked what it looked like, he said there was a bath there. He added that there was a plug in the washbasin that you could pull and that this plug was not one of those with a chain. As the police confirmed that one could not see into the bathroom from the place by the front door where S. maintained the boy had stood, and that the construction of the plug, which the boy had described quite correctly, could not possibly be observed except from inside the bathroom, S.'s statement seemed to be false. S. was questioned further on this point, but he maintained that he had not noticed whether the boy had been into the bathroom on his own. S. denied as completely absurd the idea that he should have taken the boy into the bathroom and there committed the indecent act.

The "reliability test by the psychiatrist," based on the tape-recorded interview with the boy and the conversation with his mother, resulted in a categorical opinion in favor of the boy's reliability. In the words of the expert's summary:

> The psychiatric investigation has shown Lars' statements to be reliable. He is bright and lively and of

> normally developed intelligence. In spite of his mother's statement about the child's imaginativeness, he has shown no tendency to make up stories. Nor is he in any way maliciously disposed towards certain people, and his statements are consistent. There is no doubt that the man in question stuck his penis into the boy's mouth and that the boy sucked it. . . . I therefore certify that Lars' statements, that is, what his mother reported to the police and what came out of my investigations, are highly reliable.

The security investigation of S. was entirely in his favor. He had been liked in all the places where he had been employed, and his testimonials were exclusively good ones. The Russian connection came from the fact that he was born in a part of Finland that had been ceded to Russia at the last peace treaty. S. spoke good Swedish and had been engaged for 3½ years to a Swedish girl, whom he proposed to marry as soon as his naturalization had gone through. His fiancée and her mother were convinced that he had not committed the act of which he was accused. This was the view of all those who had had anything to do with S. since he arrived in Sweden.

In spite of the good testimonials, S.'s case could only be described as particularly bad at the conclusion of the police enquiry. He had indeed denied the indecent act, but he had admitted that a boy, who might be Lars, had visited him at the time in question. He denied that he had taken the boy with him into the flat, but the enquiry showed that the boy must have passed the same way as S. into his flat, and it was impossible for a boy of Lars' age to do this on his own. S.'s statement that the boy had only stood by the door was contradicted by the boy's ability to describe details in the bathroom, which he could not have seen from the front door. This evidence was further strengthened by the fact that Lars' own house had no such conveniences as were in S.'s. It could therefore be assumed that Lars had had no previous opportunity of seeing a washbasin plug of the type he had described. Finally, a test of the child's reliability had come to the conclusion that the indecency had taken place in the way the mother had said to the police.

INVESTIGATION FROM AN INTERACTION APPROACH

In this situation, the prosecution asked for a new specialist enquiry, feeling that the case needed further elucidation. The defendant's counsel had for his part not asked for any further enquiry.

The new enquiry was conducted by the author on different principles. Instead of focusing on the personality of the child witness, he followed an approach emphasizing analysis of social interaction in elucidating the circumstances and factors of personality that had combined to produce the charge against S.

The Mother

The first phase of the enquiry was a searching interview with the boy's mother, from which the following facts emerged:

> Both parents considered that Lars was a very capable boy, advanced for his age, who took a lively interest in everything that happened around him. He had learned early to tell the truth, and, as far as his mother could judge, never lied, either to his father or mother. Besides, they had strictly forbidden him to lie. Their interest in the boy, until some two months ago their only child, was unusually strong. To show this, the mother told how the children in their district used to go round the houses and collect newspapers, which they later sold. The mother had, however, forbidden Lars to take part in such occupations unless an adult was with him. She had on a few occasions gone with him herself when he collected papers, and had then been careful to see that he bought nothing unwise with the money he earned. Both parents were very careful that he should not acquire any bad habits where money was concerned. As an example of this, the mother told how there was a dirty, badly cared for boy in the neighbourhood called Sune, who previously had played a lot with Lars. The mother had really been sorry for this Sune, who was universally unpopular in the district. Some time ago, however, she had forbidden Lars to be in Sune's company. The cause of this ban, which also included a ban on playing where Sune lived, was that she had once met the two boys together when Sune had had a whole krona (dollar). On this occasion, both boys had been lavishly provided with sweets, and when the mother questioned Sune, who was two years older than Lars, she gathered that Sune had filched the money. On one or more occasions before the incident with the window cleaner, Lars had been to Sune's house and played there, but his mother was convinced that he had not played with Sune since her prohibition. When she was asked if she knew whether Lars had ever collected newspapers on his own, she recalled that he had on one or two occasions been out and collected them together with older children. His mother did not know which houses they had visited on these occasions. When in this connection the investigator remarked that Lars was evidently quite experienced in going around ringing at the doors of strange houses to beg for newspapers, the mother conceded this without further discussion. She repeated, however, that she had forbidden him to do this by himself and that both she and the father were very particular that he obeyed their instructions.
>
> As a further example of how alert and fearless her boy was, the mother said that Lars was very much

interested in different jobs and that he often played at doing such jobs as he could observe outside. Once in town he had gone up to a policeman and asked where "Mister Policeman" had put his radio-car. His mother said that he often talked to workmen about their work.

The major part of the conversation was devoted to the circumstances in which Lars' mother had talked to him on her return from shopping. What had specially surprised her was that the boy had shown no signs of discomfort or disgust. She had rather had the impression that he wanted to stay out and play in the street, when she and her mother showed signs of disquiet and took him indoors. His mother repeatedly reproduced the boy's tone and gestures when he came out with his sensational charge. Her impression was that Lars felt he was telling them something very interesting but that he did not at all understand the significance of this action. The correct version of his words was: "Just think, mummy, I had to suck his doodle."

The mother was also surprised that the boy could not be induced to repeat the accusation later. When his father came home he had tried to talk to the boy, who would, however, give no information about what the man had done to him. Even when his mother took Lars along to the police station in the evening to make her deposition, he would not tell anything to the policewoman who interviewed him. The interview had therefore been discontinued, but next afternoon a doctor (the psychiatrist) interviewed Lars. His mother then told the doctor all she knew, and, as she had understood it, the boy then told the doctor the whole story.

Regarding the boy's open fly buttons, the mother said that both women had been excited and bewildered at the discovery. Lars' mother could not recall whether her mother had framed her question to the boy in a leading way. She knew, however, that the grandmother was already suspicious of S. before the boy was questioned.

The Police

Of all the mother's disclosures, the most interesting to the investigator was that Lars had been interviewed by the police the very same evening the deposition was made. No notes of this interview were to be found in the police report. When the Inspector concerned was asked about it, it turned out that the interview had been considered a failure because the boy could not be induced to say anything. The notes had been thrown away, and, for various reasons that could not be more exactly specified, it had not been thought necessary to mention the "unsuccessful interview" in the report. After a special request, the policewoman concerned gave an account of what had happened. In addition to what was known previously, it transpired from her account that, in talking about the flat, the boy "had been there before and got newspapers, but at that time the 'lady' was there. Questioned about the indecent assault he became noticeably irritated and the earlier contact disappeared. . . . Only at the beginning of the talk, when Lars was specially interested in the typewriter, did he sit still on a chair. Later he ran around the room and played with the office furniture."

There was no doubt that a good deal of significant data lay in this discarded interview. Points which specially emerge are the boy's habits in collecting papers and his reluctance to repeat the charge itself.

The discovery that the police report was incomplete on an important point led to further investigations concerning the information it contained. Special interest was directed at the reproduction in direct speech of Lars' charges, which prefaced the report and made S.'s case more serious. Neither the choice of words nor the content agreed at all with the description given by the mother. Further questions revealed that this piece was a summary of the impression of the officials concerned at the time of the actual deposition, not a factual statement of what Lars had actually said.

According to the police report, the boy had said of his experience, "It was so disgusting." As appears from the conversation with the mother, the boy had never uttered such words but on the contrary gave the impression that he did not react emotionally. The form of words must therefore have originated from the author of the police report, who unintentionally amplified his account in accordance with his own expectations. This well-known phenomenon is chiefly notable because the amplification in this case has crept into the report in a form exceedingly dangerous for the accused. The conclusions of the test of the boy's reliability rested among other things on this statement of his feelings of disgust.

Lars: The Accuser

The next stage in the work of investigation was to study Lars himself. His degree of intellectual development was assessed by an expert in intelligence testing, with results summarized as follows:

> At an intelligence test carried out on Saturday, June 6th, Lars scored an IQ of 123 (Terman-Merrill scale modified by Sandels). The experiment shows that it is specially easy to establish personal contact with Lars,

who gives the impression of being a child with healthy reactions. His curiosity is very great and his manner is spontaneous and marked by a certain easy self-assurance.

Lars was further examined by an expert in problems of suggestibility. (The results of this inquiry, undertaken in connection with a scrutiny of the incriminating tape-recorded interview, will be discussed later.) The author took part as an observer in some of the "suggestibility tests" and later amplified these observations with his own examination of the boy. It was easy to verify that the boy could only with difficulty be kept concentrating on the tasks he was given. He constantly slipped off his chair and roamed unhindered round the laboratory, examining the apparatus that stood within reach. Instead of answering the questions put to him, he asked his own questions about objects in the room and showed a pattern typical of many children of evasive tactics and distraction devices when in one way or another he felt unsure of himself.

The most interesting reaction came when the author during a pause in the suggestibility tests threw in the question: "Listen, Lars, do you know anyone called Sune?" Lars did not answer the question verbally. Instead he slipped off the sofa, went up to the author and gave him a crack on the jaw. It was not a real blow, but it was intended to represent a real crack on the jaw and fulfilled the function both of distraction and of symbolic demonstration of his antipathy to that particular question. Some five minutes later the question was repeated, whereupon Lars went through exactly the same performance. Since the situation seemed to be full of potential information, the question was repeated a third time but not until some ten minutes later, while other things were being dealt with. This time, the boy reacted by slipping off the sofa, so that he became invisible behind a table covered by a cloth. Down on the floor he said, drawing out the vowels: "Hoow noooosy you are."

A few days later, the author made a further personal examination of Lars. On this occasion as well, he demonstrated the distraction technique just described as soon as anything embarrassing came into the conversation.

After the investigation of Lars, the enquiry took a turn that had not been originally foreseen. So far, the author had not gone more closely into the question of S.'s possible innocence but had rather worked with the intention of studying the factual accuracy of the details of the charge. Such an aim was moreover supported by the police enquiry so very unfavorable to S.

A preliminary collation of the data so far produced by the psychological enquiry led, however, to trying out a completely new hypothesis. What suggested this idea more than anything was the boy's tendency to divert the adult's attention from whatever was embarassing him. A comparison of this tendency with the parents' rather naive idea that their boy never lied, and their overambitious accumulation of bans about his games and occupations, led one involuntarily to think of the possibility that Lars had simply used the assertion about the man's penis to divert his mother's attention from the things he then had on his conscience (he had visited forbidden territory, he had collected newspapers by himself, he had probably met Sune and perhaps also accepted money from him, he carried evidence of his crime under his arm and smeared round his mouth). Such an interpretation was further supported by his refusal to repeat the assertion that had caused so much trouble and excitement around him. The little boy, with all his wits about him, must soon have realized that his invention risked being revealed as a lie, and as lies were included in the list of forbidden things, he then made diligent use of his diversionary technique to avoid repeating the assertion.

S.: The Accused

The next task was to examine S. To begin with, S.'s employer, Andersson, was interviewed for information about S.'s conduct in connection with his arrest. In addition, the author obtained information from the police officers who took S. into custody, as well as from the Inspector who conducted the police enquiry. Then S. was brought from prison. He spent two days at the psychological laboratory, where his personality was examined. His intelligence was tested in the usual way and his life history reviewed in detail by two different specialists. His ability to reconstruct previous events was tested, and in a lengthy cross-examination his information concerning what happened on the fateful day and the day after was reviewed. The investigation of S. also included a visit to his home where he was joined by his fiancée. In summary, S. showed himself to be a normally intelligent man, kind and decent to those around him, careful about himself and very exacting about how he appeared in the eyes of others. As he was waiting for naturalization, he was particularly anxious to avoid trouble and disputes. He was industrious, reliable, and ambitious in his work. His future was well planned. He and his fiancée had jointly saved up to be able to make a home as soon as he became a Swedish citizen. For the present, they lived with the fiancée's mother in her flat.

As far as the alleged crime was concerned, S. appeared completely unable to understand it. He constantly came back to the charge without prompting and lamented with curses and tears the absurd and humiliating aspect of his situation, but in between pulled himself together and hoped that the truth would help him. He made the utmost efforts to reconstruct what he had done the day of Lars' visit, and stuck obstinately even to such details as pointed suspicion against himself. (He refused, for instance, to agree that the boy could have been into the bathroom for a moment when S. in the kitchen had his back to the hall, and similarly he maintained that he had not heard anyone come after him through the cellar passage.) All his statements which could be checked by adult witnesses were shown to be correct.

S.'s sexual experiences were normal, and his way of talking about such things was that of the normal person. He showed no signs of homosexual thinking. On the whole, he was not interested in children. He was faithful to his fiancée and fond of the people close to him.

To the author he appeared incapable of dissimulating, giving the impression of a naively straightforward and sincere person, hurt and uncomprehending in face of the charge he had incurred. He had no reproaches to make against anyone but tried to understand the reactions of those around.

Regarding his demeanor when arrested by the police, the facts were these:

> The police went to look for him at midday at his workplace, but the firm's proprietor, Andersson, could not say exactly where S. was, as his cleaning programme included at least 30-odd firms in different places in the town. The police then informed Andersson that S. was wanted for a serious offense and that it meant several years prison for anyone helping the criminal to keep out of the way. Andersson, who took offense at this, suggested that the police should come round to the firm about five P. M., when S. usually came in after finishing work. When S. arrived at the firm's office about four P. M., he was met by the information from his employer that the police wanted him for a serious offense. S. was at first more surprised than anything else but then dismissed the matter, saying that there must be some mistake. Andersson, still excited by his conversation with the police, wanted S. to confide in him, so that the protection of the firm could be arranged in time. S. however calmed him down and assured him that he had not done anything and that a mistake had certainly been made. To put an end to the uncertainty he rang up the police on his own initiative, told them his name and asked for information as to why he was wanted. He managed to learn nothing from this conversation, as attested partly by Andersson and partly by someone who knew both S. and Andersson. S. then proposed that they should sit down and have a beer while they waited for the police. About half an hour later the police arrived and arrested S. At this there was some trouble, caused by Andersson's disapproval of the way the police behaved. S. himself made no difficulty. According to the Inspector, he had some outbursts later and behaved in an unbalanced way. S. himself said that when he came to the police station he had asked for permission to shave, as he was dirty after the day's work. He had then been given a shove and told to hold his tongue. "We decide things round here, not you," someone had said to him.

Resolution of Discrepancies

The investigation of S. left the hypothesis of his innocence still more promising than when the investigation of the boy was concluded. Nevertheless, this hypothesis could still not be accepted because of several important circumstances that remained unexplained. These were: (*a*) that Lars had been able to get through the locked cellar by himself; (*b*) that Lars could describe what the washbasin plug in S.'s bathroom looked like without having been in there; and (*c*) that Lars repeated his charge at the interview with the psychiatrist the day after the incident.

When the enquiry was resumed it was now directed at the unexplained details in Lars' experiences. Together with S. and the policeman who first investigated the scene of the crime, the author visited the house where S. lived. On this occasion, S. indicated the door through which he had passed into the cellar. This door was locked. It was not, however, identical with the door indicated by Lars. S. pointed out that the door he used to go through was the most natural one to choose because it made direct contact with the staircase. From it a straight passage led to another door, which connected immediately with the staircase on which S. lived.

At the same time it was discovered that the door which the boy had previously indicated also led into the cellar and that via this door one could reach another door, which led out into the courtyard at the side of S.'s staircase. To the policeman's surprise it turned out that both these doors were unlocked on the occasion of this unannounced visit. A visit was therefore paid to the landlord's agent in the building. Asked about the closing of the cellar doors he answered: "Yes, they should always be locked."—"But they hardly ever are," added his wife, who stood beside him and evidently was less concerned with professional prestige than her husband.

After this breach in his defenses, information poured from the caretaker. He had a lot of trouble with the children in the house and children from the neighborhood, because they often played down in the cellar and carried on as they pleased. It was impossible to keep locked the doors which the children used, since people took their bicycles in that way and often forgot to lock up behind them. The agent had had particular trouble from a little hooligan called Sune Olsson. Once he had lit a fire in the cellar and another time he had damaged every bicycle that was down there. However, Sune had moved about three weeks ago, as they had managed to evict the family, who had made things very difficult, and who kept up a permanent feud with the neighbours and used foul and vulgar language.

To get an insight into Sune's living conditions, the author paid a visit to the house where he had lived. The new tenants said the flat had been badly looked after and that the previous tenants had gone out in a rush without anyone knowing where they had gone. The interesting thing about the flat was that, apart from being the other way round, it looked exactly like the one S. lived in. The bathroom looked precisely like the one the police had studied in S.'s flat, and the washbasin plug was the same. The author had no difficulty in imagining how the inquisitive little boy had run around in the Olssons' flat. The fact that he had been able to describe the plug to the police no longer seemed so extraordinary, and the association existing from the first between Sune's and S.'s homes also became more significant when the similarity became clear, because the distance between the two staircases was so great that an adult would hardly think of saying that they both lived in the same house.

Keen work by the police enabled them eventually to trace the Olsson family to a place a long way outside the town, where they had withdrawn from the demands of society and neighbors. Sune was then brought back to the family's old home, and asked to show the cellar he formerly played in. As soon as he got out of the car he went straight to the cellar entrance through which Lars claimed to have entered. In this cellar, where Sune evidently felt at home, the author talked to him about the playmates he had had. He knew at once who Lars was, and was able to describe him quite well. The last time they had met was inside in the courtyard, when Lars had some newspapers with him, said Sune. He supported this assertion by adding that he had left immediately afterwards. Asked if he had given Lars any money on that occasion, Sune gave no definite answer. It proved on the whole hard to get him to say anything if he suspected criticism behind the question. He behaved as children do who are accustomed to get the blame for everything that goes wrong: afraid, suspicious of strangers, at odds with society outside his own family.

Later the same day Sune's IQ was determined to be 92, quite sufficient for him to be able to understand the questions and answers exchanged with him.

Before Sune was taken back to his family's new retreat, his vocabulary was tested. It proved to be quite poor, but he had good knowledge of a number of vulgar expressions. He said he did not know the meaning of the word "doodle," but he knew several other names for the male sex organ. A number of hair-raising tales of how the mother had spoken to neighbors, recounted earlier by the agent, fill out the picture of the Olsson family's speech habits. There is therefore nothing against the assumption that Sune could have been the source of Lars' inspiration for the statement about how the window cleaner had acted. The fact that Lars in his version used the word "doodle" does not constitute a serious objection, because it must be assumed that Lars was intelligent enough to use such terms in speaking to his mother as were sanctioned in the family's ordinary speech.

The question of where the coin had come from was still not definitely solved. According to Lars' mother, the agent and the candy-store clerks, it often happened that Sune had money. Logically, there is therefore hardly any need to adhere to Lars' statement that he had got the money from the "window-cleaner man." He could equally well have got it from Sune.

A few more small details remained which told against S. Such a detail was the unbuttoned trousers. But it is more usual for small boys to go around with their fly buttons undone after urinating on their own than for them to do them up again. For Lars in the present situation to blame the window cleaner was quite in accordance with his other behavior.

But wasn't it odd that Lars had rung just at S.'s door? Hardly, since at the time of day when this happened, most flats in the building were likely to have been empty. The special thing about S.'s door was that there was someone behind it who could open it and satisfy the little boy's wishes.

THE INCRIMINATING INTERVIEW RE-EXAMINED

The purpose of giving such a detailed account of a single case has been to illustrate the efficacy of two different working methods. While one case does not permit a conclusive assessment of their relative merit, the present study seems in the author's opinion to indicate that the description of events given by social interaction analysis is much superior to the personality analysis method of investigating witnesses.

In this connection, we may look more closely at the tape-recorded interview by the psychiatrist, as well as because it was here that Lars repeated his charge against S. The technical deficiencies in interviewing are to be understood in the context of the inexperience of the psychiatric expert in witness psychology, and the fact that the training of a child psychiatrist in Sweden is hardly directed at this type of task.

The analysis of the interview proceeded in two phases. To begin with, Lars' reactions to the kinds of questions used in the interview were explored in the "suggestibility test" previously mentioned. Then the questions and answers of the interview were reviewed critically in order to bring out what may have lain behind the boy's answers.

The "suggestibility test" was intended to assess (*a*) the tendency of the witness to state and sometimes even to experience sense perceptions that have no objective basis; (*b*) his tendency to yield to authorities enjoying a certain prestige; and (*c*) his tendency to answer certain types of suggestion with information that the interviewer is expecting.

The first type of test showed an obvious tendency on the boy's part to accept uncritically different kinds of sense perception. The record says, "One can plainly observe how the subject was induced to state auditory perceptions after quite short intervals, although there was no objective basis for such perceptions." The same phenomenon was established with tactile perceptions.

With prestige suggestions he showed less interest. The record says, "Lars was not interested and very inattentive in this test. He cannot be considered to have received any very deep impression from the interviewer's suggestions. When he accepted the interviewer's proposals, it seemed almost to be in order to get away from the whole thing as quickly as possible."

The test with different kinds of suggestive questions was of the greatest importance for the enquiry. The questions were partly anticipatory of the type, "You do see . . ." or "Didn't you see that . . . ," partly alternative questions of the type, "Was the car blue or green?" when the car was in fact red. Finally, simple *yes* and *no* questions were tested, which also have a certain suggestive effect, as they are more often answered in the affirmative than the negative.

The record shows that Lars could be brought to assent to statements that had no basis in his own observations. He could even be induced to demonstrate with sounds and gestures things which did not exist. For example, the interviewer asked Lars if he had not seen a small pair of scissors on the experimental table. Lars said he had, although there were no scissors among the objects he had seen. When the interviewer asked the boy how big they were, Lars demonstrated the approximate size of a pair of scissors with his hands. The interviewer then asked if they were not bigger and measured out a bigger size with his hands. The willing little boy then himself measured out a bigger size with his hands and answered: "Yes, like this."

The examination of the tape-recorded interview shows that the questions put were in the main anticipatory or framed as statements to which the boy had only to say *yes* or *no*. Instead of sorting out what really had happened, the interviewer tried to get the boy to repeat his charge. The result of the interview, to summarize, was a repetition of the mother's statements put into the child's mouth. The accusation only came up when the interviewer had used both bribes and pressure.

> Dr.: But it would be nice to get straight just what happened with that man. Then you can go home with your mummy and take the aeroplane with you, and we won't have to meet again.
>
> Lars: Whose are these toys?
>
> Dr.: The toys are mine, and we've promised you that you can take an aeroplane with you. But you must tell just exactly what happened with the man, or perhaps you want another of the toys, but of course you would rather have the aeroplane, wouldn't you?

After the dialogue just quoted, Lars managed to distract the interviewer's attention, and the talk ranged for a long time over peripheral subjects. In the end the boy wanted to go to his mother.

> Dr.: But first you can surely tell me something about the man. Then you can go home with your mummy.

This attempt succeeded no better, however. After a break, when the boy talked to his mother for a while, the interview was resumed, and now the interviewer altered the situation. He proposed in fact that Lars should tell him what he had previously told his mother.

> Dr.: Now what was it you told your mummy? When mummy met you? Can't you tell me about it too?

If one accepts that the boy lied the day before and was now afraid to lie once more, the interviewer's request must have seemed inviting, for to talk about what he told his mother the day before hardly seems like a new lie from a child's point of view. In spite of this it took no less than four different requests from the interviewer before the boy, guided by the interviewer's yes and no questions, began to take up the assault story. It is striking to see how he is able to dominate the adult even in this situation. When Lars answered the question: "Did you go into any room with the man?" with "Yes, I went into their bathroom," the interviewer became more careful and avoided putting answers into the child's mouth. When the boy was thus left with no pointers as to what answers the "doctor man" wanted, he tricked the interviewer back into the old technique.

> Dr.: Did the man do anything with you or did you do anything with the man, or what did happen in the WC?
>
> Lars: *Guess!*

It is hardly surprising that after this invitation the interviewer managed to guess what Lars had told his mother the day before. The central element in the interview was thus a reproduction of what the mother told the police and the first interviewer.

On the question of the boy's first acquaintance with S. the interview contains a number of statements which depend on the fact that the boy had not been addressed by S. but had followed him on his own initiative. The following exchange shows this clearly:

> Dr.: So you stayed both up there on the stairs?
>
> Lars: (silence).
>
> Dr.: Well?
>
> Lars: No, I went in to him.

Another exchange in the same vein is the following:

> Dr.: Yes, so the man came? [that is, up to the boy and asked him into the flat in accordance with the nterviewer's presupposition].
>
> Lars: *He was there.*

The analysis of the interview clearly shows that the conclusions drawn by the interviewer go far beyond what the material legitimately supports. An experienced and skillful interviewer would never have drawn his conclusions with such assurance and would probably have produced several leads calling for further investigation; it nevertheless appears improbable that a real solution of the problem could have been achieved without an extension of the enquiry to the plane of social interaction.

WHAT HAPPENED: A RECONSTRUCTION

A man is cleaning a shop window in a suburban street. A little way off a small boy stands watching him. He would like to be a window cleaner too. The man finishes work. He collects his things and his bicycle but is stopped in the street by another man who begs for money. The little boy goes up and asks the man if he may carry his ladder. The boy often does this sort of thing. He talks to policemen and workmen. "No," says the man, "it is too heavy for you."

He is a nice man, he does not snap at anyone, children or beggars. Then he pushes his bicycle round the corner and past the next block, as far as the sweetshop, right by the house where he lives. The little boy wanders after him, as curious as ever. He sees the man put his bicycle and ladder against the wall of the shop. Then the man disappears into a house further off, where the boy used to play before mummy told him he must not go so far away.

The man has disappeared, but the boy goes into the house too. He goes through the cellar door, where children often do go in. He goes on through to the other side and out through the bicycle exit and comes into the courtyard. He has been there before too. He has been collecting newspapers in this house. His mother is shopping, and there are a lot of things that can be done without her knowing.

He goes up into the house and rings at the doors, but nobody opens. Not until he comes to the third floor is there a door that opens, and the man who opens it is the one who cleans windows. The boy asks for papers, and the nice man goes into the kitchen to fetch some. The little boy sees that this is just like where Sune lives, further off in the building. Two rooms and kitchen, with a hall and bathroom. He knows that is what Sune has, because he has been home to Sune's and played in his apartment, before mummy said he must not play with Sune. Then the man comes with the papers and gives them to him. He goes down the stairs with them under his arm. At first his thoughts are full of the window cleaner and that he has been to the man's apartment. But then he thinks that that is because he has gone with him into the house and gone up to his apartment.

When the boy comes out into the courtyard, Sune is standing there. And Sune asks where he got the papers. But when the boy tells Sune that he got them from a man, Sune tells him something old men do with children when you go to their place. And the little boy thinks that he has been to a man who does that.

Then he gets a big black penny from Sune, which Sune found at home, and he leaves Sune and goes around the house, which is ever so long, and when he comes around the other corner, he sees the man standing on his ladder cleaning the sweetshop windows.

Then he goes into the shop and spends the penny on a bit of liquorice. It buys him no more than half a bit, because the lady says there is not enough for more. He looks at the man who gave him the papers, and when he goes out of the shop he thinks about what Sune told him about old men, and then he goes home. He is not thinking of anything much any more, but he has a feeling in his stomach that he has done something he should not. Collecting papers. Being with Sune. And what Sune told him about. And then he meets mummy and granny, and they ask about the papers he got from the window-cleaner man.

"But how do you know him?" says mother anxiously.

And the little boy explains.

"I went home with him," he says.

If a man asks you in with him to collect something, no one can help that.

"And the candy," says mother, "where did you get that?"

He is not allowed to be with Sune. Perhaps it was the window cleaner who gave him the money. Old men give little children money. You go home with them, and then you get a penny.

"The window cleaner," he says.

But mother and grandmother are getting upset and hushing him up and pulling him to the stairs. There is something wrong with what he said. But he does not want to go up now, and so he gives his mother his sweet look and says:

"Just think, mummy, I had to suck his doodle."

It is a smack on the jaw for mother and grandmother, the biggest smack in the world.

But he still does not manage to stay out. Grandmother sees that his fly buttons are undone. But the window cleaner did that. Mummy talks to daddy on the telephone. The police will take the window cleaner, and he will say that Lars did not suck his doodle at all. So the boy says nothing to his daddy, when daddy asks after dinner. When he goes with mummy to the police, and the lady who has a typewriter talks to him, he does not tell her anything either.

But the police and the other grown-ups are catching the foreigner who has committed an indecent assault. In the end, a doctor comes and helps the police. He delivers his judgment after being two hours with the little boy, who besides being bright and inquisitive is not afraid of the silly grown-ups who only understand bits of his thoughts.

And society prepares to put out of harm's way someone who opened a door and gave a boy a dozen papers. It is very interesting and at the same time somehow thrilling. As drama, it is oppressive and horrifying.

It must be added that the complete reliability investigation led the prosecutor to decide that there was no case against S. After six weeks, S. left prison a free man. When he was set free no one told him why he was not to be kept in jail any longer.

Received June 10, 1957.

CRITIQUE AND NOTES

ACCURACY OF SELF-PERCEPTION AND GROUP-PERCEPTION AS RELATED TO TWO LEADERSHIP ROLES[1]

DAPHNE E. BUGENTAL AND GEORGE F. J. LEHNER

University of California, Los Angeles

IN the extensive literature on leadership, there is a frequent tendency to treat leadership as a unitary concept that may be measured by a wide variety of sociometric questions, ratings by peers or supervisors, observational techniques, etc. Consideration is seldom given to the possibility that different types of leadership or status are measured when different criteria for leadership are used. Gibb (4), an exception, found that persons who were socio-telic or psyche-telic choices of other group members were different from individuals found to be leaders by observational techniques. He also found that individuals viewed by the group as having "influence" more closely compared with the leaders selected by observational techniques than with the socio-telic or psyche-telic choices. Bales (1) also gave consideration to the possibility of different leadership roles. In observing groups over time, he found that the role of "best-liked" group member became quite distinct from the "leader" role.

Two types of status or leadership roles appear to be distinguishable. The first is the role of the true "leader" whose primary function is the guidance of the group towards their goals. The second "leadership" role is that of the popular group member, which is roughly equivalent to Bales' "best-liked" role and Gibb's "psyche-telic" and "socio-telic" choices. The popular person is the center of positive affect within the group. However, hostility may be directed toward the individual enacting the "leader" role. The popular person does not assume responsibility for reaching group goals, and he interacts on an emotional or "social" level rather than the task-oriented level of the leader. These two roles may overlap and the same person play both roles, but they are distinct roles. Probably they are more dichotomous in largely task-oriented groups.

In view of the characteristics of these two roles, the popular member would appear to be in the better position to perceive accurately other group members and the group as a whole; that is, he should be more likely to know how well other group members are liked or how they feel about significant group issues. His greater "insight" is predicted from the sociocentrality of his role within the group.

Two studies concerned with the ability of leaders to estimate group opinion found discrepant findings. Chowdhry and Newcomb (2) found that "leaders" (defined as individuals sociometrically chosen on questions measuring both popularity and leadership variables) had better insight than isolates or nonleaders on group-relevant issues. Hites and Campbell (7), on the other hand, found that "leaders" (defined as elected or appointed leaders) did not differ significantly from nonleaders in ability to estimate group opinion on group-relevant issues. These differences in findings, to the extent that the reported findings are free of artifactual components, may have been influenced by the two ways of measuring leadership.

Several studies (3, 5, 8, 9) have been concerned with the insight of leaders (most frequently measured by popularity variables) into self and others on social and personality variables. In general, leaders were found to have better insight than nonleaders into others. The evidence for greater self-insight was not strong, although there was some evidence for greater self-insight with higher status.

The intended contribution of the present study is the prediction that it is the popular person, not the leader, who has more accurate perception of the group structure and of group members. It is also predicted that the popular person has more accurate self-perception than does the leader. Our main interest in testing this hypothesis is to determine whether the popular group member is generally more "insightful" or if his accuracy of perception is related specifically to his role within the group. If popular individuals have both more accurate self-perception and more accurate group-perception than do leaders, they would appear to be more "insightful" in general. If they have more accurate perception of the group but not of themselves, their "insight" would be seen as primarily associated with their position within the group. Our specific hypotheses are:

Hypothesis 1. Popular group members are more accurate than leaders in their perception of status relationships within the group.

Hypothesis 2. Popular group members are more accurate than leaders in their perception of significant group dimensions.

---

[1] Appreciation is expressed to R. T. Morris for his suggestions in the conducting of the study and for the use of his students as *S*s.

Hypothesis 3. Popular group members have more accurate self-perception than leaders on personality variables.

METHOD

Members of a coeducational upper division sociology class served as subjects (*S*s). The class ($N = 48$) was randomly divided into five groups of 8 to 11 members each with the purpose of completing a class project. These groups had been meeting for a semester at the time of testing.

In order to distinguish between "leaders" and "popular" group members, we drew on Bales' (1) discussion of leader roles to devise a questionnaire consisting of a series of descriptions of members, part of which related to the "leader" role and part to the "popularity" role. Examples of "leader" descriptions are:

1. His (her) suggestions for group action are usually accepted and followed by the group.
2. He (she) assumes responsibility for getting group work done and group problems solved.

Examples of "popularity" items are:

1. He (she) is well-liked by almost all, or all group members.
2. Group members seek him (her) out as a pleasant person to be around outside the group.

Each person in a group rated all other members of his group on eighteen items of this type on a five-point scale. *S*s were instructed to check "A" if the description was "definitely true" of the member they were rating, "B" if it was "mostly true," "C" if it was to an equal degree true and false or if they were unable to say, "D" if it was "mostly false," and "E" if it was "definitely false." Ratings were thus obtained on all *S*s on both leadership and popularity variables.

All *S*s were then ranked for mean rating received on leadership and on popularity. *S*s in the top 25 per cent on leadership ratings were defined as "leaders"; those in the top 25 per cent on popularity ratings, as "populars." As we wished to distinguish between the "leader" role and the "popular" role, overlapping cases ($N = 6$) were eliminated. All further comparisons were based on seven leaders and five popular members.[2]

To determine whether popular group members were more accurate than leaders in their perception of group status relationships, we again used our group member descriptions questionnaire. The closeness of agreement between a member's descriptions and those of the group as a whole provided a measure of consensual accuracy, in terms of a mean error score based upon the differences between a person's ratings of other group members and the group's ratings of these same members. Insight into popularity and into leadership were measured separately.

To obtain measures of accuracy of perception of significant group dimensions, we used the Hemphill Group Dimensions Descriptions Questionnaire (6). The Hemphill questionnaire measures 13 group dimensions, such as viscidity, hedonic tone, autonomy, etc.; one of these dimensions (permeability) was omitted as inappropriate. The questionnaire requires members to rate their group on a five-point rating scale on a series of statements such as, "The group has well understood but unwritten rules concerning member conduct." Again we measured accuracy consensually, so that an individual who was very discrepant from the rest of the group in his ratings of the group had a high mean error score, and one who agreed closely with the group as a whole had a low mean error score.

TABLE 1

MEDIAN ERROR SCORES[a] ON RATINGS OF GROUP MEMBERS AND GROUP DIMENSIONS

Variables Rated	"Populars" ($N = 5$)	"Leaders" ($N = 7$)	U
Popularity of others	.59	.87	7*
Leadership of others	.85	.77	16
Hemphill group dimensions	.56	.87	3**

[a] Error scores represent mean standard scores.

* Significant at the .05 level of confidence.

** Significant at the .01 level of confidence.

Accuracy of self-perception on personality variables was measured with the Guilford-Zimmerman Temperament Survey and a self-rating test in which *S*s rated themselves on the factors measured by the Guilford-Zimmerman. T scores for the leaders and populars based on distributions of self-ratings on each factor for the total group were compared with T scores received on the same factors as measured by the Guilford-Zimmerman, and difference or error scores were computed.

RESULTS

In comparing leaders with popular group members for accuracy of perception (indexed by error scores) we used a one-tailed nonparametric measure, the Mann-Whitney U test. In terms of our hypotheses, our results were as follows:

Hypothesis 1. Popular members, as predicted, were more accurate than leaders in their perception of the popularity of others. ($U = 7, p < .05$.) However, there was no difference between leaders and popular group members in accuracy of perception of the leadership status of others. ($U = 16$, n.s.)

Hypothesis 2. As predicted, popular members were more accurate than leaders in their perception of significant group dimensions. ($U = 3$, $p < .05$.)

Hypothesis 3. Popular group members did not differ from leaders in accuracy of perception of self on personality variables. ($U = 17$, n.s.)

DISCUSSION

In the light of these results it appears that individuals enacting the role of the popular member have more accurate perception of the popularity status of others and of significant group dimensions than do members enacting the leader role. It is difficult, however, to say whether the popular person achieves this role because of his more accurate perception of the group, or whether he has more accurate perception of the group because of his position within the group. Both relationships are plausible.

---

[2] Two other populars had to be eliminated because of incomplete returns.

Apparently, the popularity role does not include more accurate perception of the leadership status of others. The well-liked person appears to have superior accuracy of perception only with respect to factors specifically related to his role. His understanding of the group relates to how well other members are liked, how members feel about the group, etc.—but does not include a superior understanding of who is directing the group towards their goals. Additional evidence for the belief that the popular person's more accurate perception is specifically tied to areas related to his role is given by our finding that there is no difference between the popular person and the leader in accuracy of self-perception. The "insight" of the popular person is not general, then, but is confined to variables associated with the sociocentrality of his role.

SUMMARY

"Popular" group members were compared with "leaders" of five small groups for accuracy of perception of other group members on popularity and leadership variables, accuracy of perception of the group as a whole on 12 group dimensions, and accuracy of perception of themselves on personality variables. "Popular" members were found to have significantly more accurate perception of others on popularity variables, and of the group as a whole on the group dimensions measured. No significant differences were found between "popular" members and "leaders" in accuracy of self-perception on personality variables or in accuracy of perception of others on leadership variables.

REFERENCES

1. BALES, R. F. The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. A. Shils (Eds.), *Working papers in the theory of action*. Glencoe, Ill.: Free Press, 1953. Pp. 111–161.
2. CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.
3. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18.
4. GIBB, C. A. The sociometry of leadership in temporary groups. *Sociometry*, 1950, **13**, 226–243.
5. GREEN, G. H. Insight and group adjustment. *J. abnorm. soc. Psychol.*, 1948, **43**, 49–61.
6. HEMPHILL, J. K. *Group dimensions*. Columbus, Ohio: Bur. of Business Res., 1956.
7. HITES, R. W., & CAMPBELL, D. T. A test of the ability of fraternity leaders to estimate group opinion. *J. soc. Psychol.*, 1950, **32**, 95–100.
8. LEMANN, T. B., & SOLOMON, R. L. Group characteristics as revealed in sociometric patterns and personality ratings. *Sociometry*, 1952, **15**, 7–90.
9. NORMAN, R. D. The interrelationships among acceptance-rejection, self-other identity, insight into self, and realistic perception of others. *J. soc. Psychol.*, 1953, **37**, 205–235.

Received April 29, 1957.

BEHAVIOR AND GRAMMATICAL STYLE

LEONARD W. DOOB
Yale University

Do people express themselves in such characteristic ways that a relatively simple analysis of their grammatical style can yield useful clinical insights? Style as well as content are said to spring from enduring predispositions: "the frequency with which a person employs a certain grammatical category such as nouns, verbs, adjectives and the like has been shown to be related to various personality characteristics" (8, pp. 248–249). The evidence cited after that assertion and existing evidence in general are spotty. By an ex-post-facto analysis, for example, the behavior of two *S*s has been linked to their prose style (9). Or the grammar employed by various types of abnormal patients has been partially differentiated (1) or compared with normal *S*s (3). A connection has been sought between a single category like verb-adjective ratio and a broad behavioral category like emotional stability (2).

METHOD

To explore a wide variety of possible relations between behavior and style, *S*s whose behavior had already been experimentally and heterogeneously sampled by Linton (5) were used.[1] A total of 28 measures was obtained from 49 male students at Brooklyn College. The measures are grouped and listed in Table 1.

---

[1] Keen appreciation is expressed to Harriet B. Linton for placing the behavioral data at my disposal, for collecting the essays, and for numerous suggestions.

TABLE 1

BEHAVIORAL MEASURES

Perceptual Situations

1. (−) Ability to locate Embedded Figures
2. (−) Ability to respond accurately in Tilting-Room-Tilting-Chair Test*
3. (−) Factorial ability to handle object in field under distracting conditions*
4. (−) Factorial ability to handle field as a whole

F-Scale and Other Questionnaires

5. (+) Conformity to convention*
6. (+) Submissiveness to parents*
7. (+) Submissiveness to authoritative figures
8. (−) Recognized self-assertiveness*
9. (+) Disguised aggressiveness

Autokinetic Situation

10. (−) General ability to adapt to the situation
11. (+) Report of relatively great movement
12. (−) Relatively fast reaction time*
13. (+) Influenced by *E*'s confederate
14. (+) Efforts to counteract *E*'s confederate*
15. (+) Generally impressed by *E*'s confederate

Verbal Communication

16. (+) Change in advocated direction

Syllogisms

17. (−) Ability to reason correctly concerning neutral items*
18. (+) Susceptibility to "atmosphere" effect

Rorschach

19. (−) High Total R
20. (+) High Total T
21. (−) High Total M*
22. (+) High Hd:H
23. (+) High F%
24. (+) High Total C
25. (−) High M:C
26. (+) High (C + CF): FC
27. (−) High Total W and/or P
28. (−) High "mammals":"non-mammals" A

TABLE 2

PROSE CATEGORIES

1. (+) Number of words*
2. (−) Words per sentence
3. (−) Variability of sentence length
4. (−) "Noun-qualifiers" (adjectives, participles, dependent clauses) per 100 words
5. (−) "Noun-qualifiers" per sentence*
6. (−) "Adverbs" (excluding negatives) per 100 words*
7. (−) "Adverbs" per sentence
8. (−) "Adverbs" per 100 "noun-qualifiers"*
9. (−) "Pronouns" (excluding ego references, including adjectives) per 100 words*
10. (−) "Pronouns" per sentence*
11. (−) "Pronouns" per 100 "noun-qualifiers"*
12. (−) "Pronouns" per 100 verbs*
13. (+) "Ego pronouns" (including adjectives) per 100 words
14. (+) "Ego pronouns" per sentence
15. (+) "Ego pronouns" per 100 "pronouns"
16. (−) Verbs per 100 words
17. (−) Verbs per sentence
18. (+) Verbs per 100 "noun-qualifiers"
19. (+) Forms of "to be" per 100 "principal verbs" (carrying the sentence's principal action, usually of the independent clause)
20. (−) Forms of "to have" (not as auxiliary) per 100 "principal verbs"*
21. (−) "Principal cognitive verbs" (involving internal reactions) per 100 "principal verbs"*
22. (−) "Principal conative verbs" (involving overt action) per 100 "principal verbs"*
23. (+) "Principal verbs" in future tense per 100 "principal verbs"
24. (+) "Principal verbs" in passive mode per 100 "principal verbs"*
25. (+) Concrete or specific "principal subjects" (the subject of the sentence's "principal verb") per 100 "principal subjects"
26. (−) Negative sentences per 100 sentences
27. (−) Sentences with "thought qualification" (clauses, phrases, adverbs, subjunctive mode) per 100 sentences.

Details concerning the "perceptual situations" can be found in Witkin *et al.* (10), concerning the other situations in Linton (5, 6, 7). Many of the measures have been found by these investigators to be significantly interrelated.

Samples of the *S*s' prose style were essays in which they described their choice of a major subject in college and their proposed career. The prose was analyzed by means of 27 categories which are relatively devoid of content and which were chosen either because they could be easily applied to any communication or because they had been used in previous studies. They are listed in Table 2. Those defined slightly differently from conventional grammar are enclosed in quotation marks and, upon their first appearance in the table, they are accompanied by an operational definition in parentheses. The reliability of the categories—tested by sampling agreement between two independent coders—is high, modally in the .90's.

RESULTS

Linton's behavioral data have been dichotomized at psychologically meaningful points whenever possible. The cut-off points for the grammatical categories have been arbitrarily established to achieve approximately equal numbers in each division. The 28 behavioral measures have then been paired with the 27 grammatical categories. A careful inspection of the chi squares which have been calculated or estimated for the resulting 756 2 x 2 tables reveals that the number of significant differences at the .05 level with a two-tail test does not exceed chance expectancy and that the few significant differences do not form useful or meaningful patterns. More refined statistical analysis scarcely seems worthwhile.

One additional analysis, however, has been made by postulating enduring dispositions which may affect both behavior and grammatical style. A freely and loosely modified version of Witkin's and Linton's psychological dichotomy has been employed, but for convenience Witkin's concepts are borrowed: a relative tendency to be *passive* ("associated with field-dependent perceptual performance") versus a relative tendency to be *active* ("associated with independent or analytical perceptual performance") (10, p. 467). In Table 1, the sign in parentheses before each measure provides a rough summary of Linton's findings: a plus sign indicates that those with passive rather than active tendencies demonstrated the behavior, a minus sign signifies the reverse.

In Table 2, the signs in parentheses suggest whether those with passive tendencies (+) or with active ones (−) can be expected to produce a *higher* frequency of the prose category. Almost all of these predictions have had to be based upon arbitrary assumptions and intuitions rather than previous research. "Passive" *S*s, it has been presumed, would cooperate more fully by writing more words [#1] and, since the task involved in part their future career, by using the future tense more frequently [#23]. They might be expected to refer to themselves fairly often [#'s 13–15], to be more interested in general action than in specifically changing the environment [#'s 18, 19], to abstract less [#25], and of course to write in the passive rather than the active mode [#24]. In contrast, "active" *S*s, it was supposed, would write more complicated prose as judged by sentence length [#2], sentence variability [#3], and the rather pronounced use of "pronouns" [#'s 9–12]. It might be anticipated, too, that they would seek to control their environment and themselves by modifying and qualifying the former [#'s 4–8, 26, 27]; and that they would use more verbs in general [#'s 16, 17] and especially those indicating activities of possessing, feeling or thinking, and overt movement [#'s 20–22].

Even though some of the measures are not independent, a simple sign test has been applied to the predictions in both tables. When the *S*s are divided into plus and minus groups on the basis a single category in one table, the plus group should contain a higher proportion of plus scores than the minus group from all or almost all the categories of the other table. The "ability to locate Embedded Figures," for example, tends to be possessed by "active" *S*s (Category 1, Table 1, is designated minus). Those *S*s demonstrating superior ability in this respect, therefore, should write fewer words since "number of words" is presumed to be associated with "passive" *S*s (Category 1, Table 2, is designated plus); they should write longer sentences (Category 2, Table 2, is minus); and so on, through the 27 categories of Table 2. Thus, each behavioral category can be viewed as being tested 27 times by means of the grammatical categories; and each grammatical category 28 times by means of the behavioral categories. Items in Tables 1 and 2 which are followed by an asterisk show significant trends at the .05 level, one-tail test. In Table 1 it can be seen that 9 of the behavioral measures show such a trend; in Table 2 the number of significant grammatical categories is 12.

DISCUSSION AND CONCLUSION

The results from applying the sign test to the grammatical categories can be briefly summarized. In contrast with those revealing a relative tendency to be passive in the behavioral situations the *S*s revealing a relative tendency to be active are inclined (*a*) to qualify statements by means of "adverbs" and "noun-qualifiers"; (*b*) to use "pronouns"; (*c*) to use verbs indicating action, possession, and internal reactions; and (*d*) to be comparatively less verbose. It is noteworthy that some parts of that summary contradict and others confirm previous studies. Here, for example, the verb-adjective quotient (Category 18, Table 2) is not discriminating, but the passive-mode index (Category 24) is; whereas apparently Balken and Masserman in their frequently cited study (1) obtained reverse results when they compared three classes of neurotics and psychotics. In both studies, however, number of words is a significant measure.

The failure of the present study to reveal more numerous significant relations per se might be variously explained. Faulty sampling may be involved: the *S*s may be too homogeneous, or Linton's behavioral measures or the pieces of prose may not have adequately reflected their enduring dispositions. In spite of the large number of indices, the grammar of the prose may not have been validly apprehended. Possibly scaling techniques are needed; those which have been here attempted—by various combinations of the grammatical categories—have not been fruitful.

Resorting to the sign test, nevertheless, produced trends which suggest that grammatical style has at least some tenuous connection with personality. That connection, however, is not specific enough to justify the statement of McClelland and his associates which was quoted above and which they themselves explicitly modify. For it varies with (*a*) the kind of behavior being observed or measured; (*b*) the categories being employed to analyze the grammatical style; and (*c*) as Goldman-Eisler (4) has shown for psychiatric patients being interviewed by different psychiatrists, the kind of situation in which the verbal habits are sampled. In addition, successful analyses of grammatical style cannot be made in the course of an interview or through any rapid device; the present categories, even though selected in part because they appear simple, require tedious counting and statistical analysis. For the moment, therefore, the following practical conclusion seems warranted: the analysis of grammatical style by itself is not a useful or feasible clinical instrument.

REFERENCES

1. BALKEN, E. R., & MASSERMAN, J. H. The language of phantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state, and obsessive-compulsive neuroses. *J. Psychol.*, 1940, **10**, 75–86.
2. BUSEMANN, A. *Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik*. Jena: Gustav Fischer, 1925.
3. FAIRBANKS, H. Studies in language behavior: II. The quantitative differentiation of samples of spoken language. *Psychol. Monogr.*, 1944, **56**, 19–38.
4. GOLDMAN-EISLER, F. A study of individual differences and of interaction in the behaviour of some aspects of language in interviews. *J. ment. Sci.*, 1954, **100**, 177–197.
5. LINTON, H. B. Relations between mode of perception and the tendency to conform. Unpublished doctoral dissertation, Yale Univer., 1952.
6. LINTON, H. B. Rorschach correlates of response to suggestion. *J. abnorm. soc. Psychol.*, 1954, **49**, 75–83.
7. LINTON, H. B. Autokinetic judgments as a measure of influence. *J. abnorm. soc. Psychol.*, 1954, **49**, 464–466.
8. MCCLELLAND, D. C., ATKINSON, J. W., *et al. The achievement motive*. New York: Appleton-Century-Crofts, 1953.
9. SANFORD, F. H. Speech and personality: a comparative case study. *Character & Pers.*, 1942, **10**, 169–198.
10. WITKIN, H. A., LEWIS, H. B., *et al. Personality through perception*. New York: Harper, 1954.

Received May 9, 1957.

ACHIEVEMENT STRIVING, SOCIAL CLASS, AND TEST ANXIETY[1]

MARTIN L. HOFFMAN
Merrill-Palmer School

SPIRO B. MITSOS
Evansville State Hospital

ROLAND E. PROTZ
Beatty Memorial Hospital

NEED Achievement scores of middle-class, as compared to working-class, adolescents have been found by Douvan (1) and Rosen (3) to be high under ordinary test conditions; and by Douvan (1) to remain at the same level even when a desired prize is offered for superior performance. Thus, achievement striving in the middle class appears to be highly internalized and relatively unaffected by external reward.

On the assumption that achievement striving relates positively to performance, it was hypothesized that middle-class performance would also be relatively unaffected by external reward; more specifically, that compared to working-class *S*s, middle-class *S*s would show little improvement in test performance under financial incentive. Two studies were done, one using a simple motor test (Study A) and the other, a written intelligence test (Study B).

METHOD

Social Class

The details of the class measure are presented by Douvan (1). In brief, *S*s designated as working class were those who identified with the working class and whose fathers did manual work; those designated as middle class identified with the middle class and their fathers' occupations largely involved dealing with people.

Study A

The procedure of Study A followed that used by Douvan (1). Students in their regular high school classrooms performed a simple number-tracing task under a time limit that guaranteed that no one would finish. In five classes the test was introduced simply as a measure of "visual-motor coordination, how well your hands and eyes work together." For six other classes it was added that in order to encourage the students to try their hardest, a five-dollar reward would be offered to every one scoring above "a certain point." In order to minimize competitive motivation, the possibility of everyone's getting the reward was emphasized.

The *S*s in Study A were high school juniors and seniors in a medium sized midwestern city (population about 80,000). From a pool of 271 students, 122 were selected who clearly fit the middle-class criteria and 37 who clearly fit the working-class criteria. It turned out that 68 middle-class and 24 working-class *S*s had taken the test under the financial reward condition while 54 middle-class and 13 working-class *S*s had performed under the no-reward condition.

Study B

In Study B, a before-after design was used. Alternate forms of the Otis Quick Scoring Mental Ability Test were given the same *S*s under the two conditions. The setting in each case was a large assembly hall in which all students took the test together. Gamma Test Form C was administered first. The procedure given in the Manual of Directions was followed except that the students were allowed to work for only ten minutes. The instructions to the students stated that the test measured one's "ability to think." Gamma Test Form C was given one week later with no advance notice that there was going to be a retest. No information was provided regarding previous level of performance. The instructions included mention that the test was similar to the one taken previously, that we were now interested in how much they could improve their scores, and that five dollars would be given to "those of you who do a good deal better on the test than they did a week ago." To minimize competitive motivation, it was added that "all sizeable improvements would be rewarded."

---

[1] Part of this paper was read at the American Psychological Association meetings in 1957.

A questionnaire asking each *S* to estimate (*a*) the extent of his desire for a five-dollar prize and (*b*) how hard he would work for this sum, was used to provide an index of conscious motivation aroused by the reward offer.

The Otis was scored in the standard manner, i.e., adding the number of items answered correctly. To take into account changes in the number of errors, the tests were also scored in terms of right minus wrong.

The *S*s in Study B were high school seniors in a small midwestern city (population about 15,000), none of whom had ever taken the Otis before. From a pool of 150 students, 47 middle-class and 27 working-class *S*s were selected using our dual criterion.

RESULTS AND DISCUSSION

Study A

The data for Study A (Table 1) confirm the hypothesis that the performance differential under the two test conditions would be smaller for the middle class than for the working class. The working class improved significantly in the reward condition while middle-class scores declined slightly.

Study B

In Study B, with either total-correct or right-minus-wrong scoring of the Otis, both the middle- and working-class *S*s obtained significantly higher scores in the reward than in the non-reward condition (Table 2). The middle class showed less improvement, as expected, but the difference is not significant ($p < .3$ with total-correct scoring and $p < .2$ with right-minus-wrong scoring). The hypothesis of less improvement in the middle class receives statistical support, however, when we disregard the amount of improvement and consider only the percentage of *S*s who show any improvement at all. Thus, with total-correct scoring, 85.2 per cent of the working-class and 61.7 per cent of the middle-class *S*s showed an improvement ($p < .02$). With right-minus-wrong scoring the percentage of working- and middle-class *S*s showing improvement are 77.8 and 55.3 respectively ($p < .03$).

TABLE 1

Middle-Class and Working-Class Performance on a Motor Task Under Two Reward Conditions

	Material Reward				
	Absent		Present		
Class	*N*	Mean	*N*	Mean	*p*
Middle	54	20.11	68	19.87	—
Working	13	16.00	24	20.00	.01

TABLE 2

Middle-Class and Working-Class Improvement from No-Reward to Reward Condition

Class	*N*	Number of Items Correct			Right Minus Wrong		
		No Reward	Reward	*t*	No Reward	Reward	*t*
Middle	47	29.7	32.4	3.84***	23.8	25.4	1.83*
Working	27	26.2	29.5	5.35***	17.1	19.9	3.52***

* $p < .05$.
*** $p < .001$.

TABLE 3

Percentage of Middle-Class and Working-Class *S*s with Low, Middle, and High Improvement

Class	Low[a]	Middle[a]	High[a]
Total correct			
Middle	46.8	21.3	31.9
Working	25.9	48.2	25.9
Right minus wrong			
Middle	46.8	19.2	34.0
Working	18.5	55.5	25.9

[a] Low, middle, and high cut-off points obtained by trisecting the total distribution based on all 150 *S*s in the initial pool.

Results of the questionnaire on the desirability of the five-dollar reward indicate, as also reported by Douvan (1), that there was no class difference in conscious desire for the reward. Apparently the reward was large enough to elicit a high degree of motivation from the *S*s regardless of class position. Therefore, the greater number of working-class *S*s showing improvement can not be attributed simply to the fact that money is scarcer in their environment.

Achievement striving and test anxiety. A greater percentage of middle-class than working-class *S*s obtained *lower* scores on the second testing when the reward was offered than on the first testing when there was no reward—25.5 per cent compared to 11.1 per cent with total-correct scoring

($p < .1$), and 38.3 per cent compared to 14.8 per cent with right-minus-wrong scoring ($p < .02$). These *S*s did more poorly in the reward condition even though (*a*) they were as motivated as the others to obtain the five dollar prize, as indicated by their questionnaire responses, (*b*) they had as much room for improvement, i.e., their initial scores were comparable to the initial scores of the others, and (*c*) they should have shown at least some improvement due to practice effect, since the two test forms were given only a week apart and contained items of highly similar content and level of difficulty. It therefore seems necessary to conclude, in line with recent test anxiety research, that the *S*s whose scores went down actually increased their strivings in response to the reward offer, but in so doing touched off efficiency-reducing emotional responses such as fears and expectations of failure.

Increased striving can also be inferred on the part of most of the remaining middle-class *S*s, since all but four of them made some improvement, especially since the middle-class improvers actually improved to a greater extent than the working-class improvers ($p < .1$ for total correct and $p < .2$ for right-minus-wrong scoring).

This analysis suggests that (*a*) most middle-class *S*s, contrary to the finding by Douvan mentioned earlier, increase their achievement strivings under external reward conditions and (*b*) the effect of such an increase in striving in the middle class is either to improve performance or to impair it. It would follow that the distribution of improvement scores for the middle class should be bimodal and highly variable. Both these expectations are borne out in our data. The middle-class distribution shows tendencies toward bimodality (Table 3) and it is more variable than the working-class distribution (F ratio significant at the .05 level with either total-correct or right-minus-wrong scoring).

McClelland (2) suggests that fantasy verbalization as well as test performance is depressed by anxiety about achievement and failure, and that as a result an S who is high on achievement striving may obtain spuriously low scores on his projective need Achievement measure if he is also high on anxiety. If such anxiety about achievement and failure can be aroused, or perhaps intensified, by an increase in striving induced by external reward, and if this occurs mainly in the middle class as the present study suggests, a number of middle-class *S*s should be expected to obtain lower need Achievement scores in reward than in non-reward conditions. This reason could explain Douvan's (1) finding that the mean need Achievement score for the middle class does not rise in the reward condition.

CONCLUSIONS

Both studies support the hypothesis that middle-class test performance is more highly generalized and less subject to improvement through increased striving for material reward than working-class performance. Study B, with before-after data, suggests further that although middle-class *S*s may have achievement strivings that are relatively highly internalized and thus generalized to situations which have no external reward, they are like the working-class *S*s in that these strivings *do* increase when a reward is offered. *The main difference between the two groups seems to be that while working-class striving and performance tend to rise uniformly in response to reward stimuli, in the middle-class reward-induced increases in striving may either raise the level of performance or touch off anxiety responses that lower it.* The source of this anxiety is a problem for further research. The literature on social class would suggest that, whether or not a prize were offered, the attempt to improve one's performance on a task designated as measuring the "ability to think" might readily engage the self-esteem needs of many middle-class *S*s; moreover, for some *S*s, attempting to improve may have this effect even with less ego-involving tasks. The resulting anxiety, which might be called "improvement anxiety" to distinguish it from other kinds of test anxiety, could have resulted in the performance decrements found here. Another possible explanation of these decrements is that the offer of a prize produced guilt feelings in those *S*s for whom competition with siblings and peers for adult rewards is synonomous with the expression of unacceptable hostile impulses. Such *S*s might have felt competitive despite the announcement by the adult that it was possible for everyone to win the prize. Doing poorly under these conditions could function as a guilt-avoiding mechanism.

REFERENCES

1. DOUVAN, E. Social status and success strivings. *J. abnorm. soc. Psychol.*, 1956, **52**, 219–224.
2. MCCLELLAND, D. C. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
3. ROSEN, B. C. The achievement syndrome. *Amer. soc. Rev.*, 1956, **21**, 203–211.

Received June 14, 1957.

BUREAU OF EDNL. & PSYL. RESE

How does our society react to the problems of mental illness? READ

SOCIAL CLASS AND MENTAL ILLNESS

A Community Study

By AUGUST B. HOLLINGSHEAD *and* FREDRICK C. REDLICH, *both of Yale University*. This is the first book to deal with the interrelationships between social class and mental illness in an urbanized community—in this case, New Haven, Conn. The authors have examined the social structure of the community, the psychiatric patients in treatment, the institutions where they are cared for, and the psychiatrists who treat them. They find the New Haven community characterized by a distinct class structure. Each class exhibits definite types of mental illness. Furthermore, each class reacts to the presence of mental illness in its members in different ways, and the treatment of psychiatric patients within the various classes differs accordingly.

Successive chapters tell the story of how members of the community become patients, how they and their families respond to psychiatric intervention, and how social class affects both therapists and patients. In the final chapters, the authors offer recommendations on what our society could do, if it wants to, about improving socially determined shortcomings of psychiatric practice.

1958. *442 pages.* *$7.50*

Send for an examination copy

JOHN WILEY & SONS, Inc. **440 Fourth Avenue, New York 16, N. Y.**

Important Evergreen Psychology Books $1 and up

The same fine soft-cover editions used in college psychology courses throughout the country. Now at your bookstore.

Masochism in Modern Man by *Theodor Reik*. $1.95

A Study of Interpersonal Relations. New Contributions to Psychiatry, edited by *Patrick Mullahy*. $2.45

Psychoanalysis: Evolution and Development. A Review of Theory and Therapy, by *Clara Thompson* and *Patrick Mullahy*. $1.75

Black Anger. A unique psychoanalytic biography of a native African medicine man, by *Wulf Sachs*. $1.75

Education and Human Relations by *Ashley Montagu*. $1.45

The Forgotten Language. An Introduction to the Understanding of Dreams, Fairytales and Myths, by *Erich Fromm*. $1.75

Rebel Without a Cause. The hypnoanalysis of a criminal psychopath, by *Robert M. Lindner*. $1.75

The Biosocial Nature of Man by *Ashley Montagu*. $1.25

Listening with the Third Ear. The Inner Experience of a Psychoanalyst by *Theodor Reik*. $1.95

Democracy and Dictatorship. An incisive study of the psychological and social conditions that produce each, by *Zevedei Barbu*. $1.45

Jung's Psychology and Its Social Meaning by *Ira Progoff*. $1.75

Oedipus: Myth and Complex. A review of all major psychoanalytic theories by *Patrick Mullahy*. $1.95

Specialized Techniques in Psychotherapy. Edited by *Gustav Bychowski, M.D.* and *J. Louise Despert, M.D.* $1.95

The Marquis de Sade. An essay by *Simone de Beauvoir*, with selections from de Sade's writings. $1.75

SEND FOR FREE DESCRIPTIVE CATALOG

grove press
795 Broadway, New York 3